TESIS DOCTORAL

ESTUDIOS ESTRUCTURALES Y COMPUTACIONALES DE AMILOIDES Y PLEGAMIENTOS NOCIVOS EN BIOMOLÉCULAS

Miguel Ángel Mompeán García

Madrid, septiembre de 2015

STRUCTURAL AND COMPUTATIONAL STUDIES OF AMYLOIDS AND NOXIOUS FOLDS IN BIOMOLECULES

by

Miguel Ángel Mompeán García

Submitted on September, 2015 to the Department of Condensed Matter Physics in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biophysics at the Universidad Autónoma de Madrid

Supervisor: Douglas V. Laurents

Instituto de Química Física "Rocasolano". Grupo de Espectroscopía de RMN de Ácidos Nucleicos.

Supervisor: Carlos González Ibáñez

Instituto de Química Física "Rocasolano". Grupo de Espectroscopía de RMN de Ácidos Nucleicos.

Tutor: Juan Salvador Jiménez

Universidad Autónoma de Madrid. Facultad de Ciencias. Departamento de Química Física Aplicada.

Programa de doctorado de Biofísica
Facultad de Química
Universidad Autónoma de Madrid

A mis padres y a mi *Fratello*

AGRADECIMIENTOS

Gracias es la primera palabra que pide paso cuando hablo de Douglas. Por tu incansable paciencia, tu generosidad sin límite y todo lo que me has enseñado, gracias Doug. Maestro, compañero y espejo donde mirarme. He descubierto una gran pasión en la investigación gracias a ti. Cómo han volado estos inolvidables años y cuánto he disfrutado.

Gracias, cómo no, a Carlos. Tus consejos son motor de este barco a remo.

Gracias Marta. Fuente de sabiduría para toda sed de duda y sendero recto para tortuosos caminos.

Jorge, M. Ángeles, Irene, José, Manuel, Treviño. Este apartado tiene un lugar guardado para vosotros. Compañeros de ciencia y cafeína, gracias por cada minuto.

Gracias a David. Más que un compañero. Un hermano mayor y excelente cocinero. Me sirve la ciencia en su punto. Siempre que puedo me escapo y disfruto ,como un niño, en su cocina.

En la memoria guardo a Alfredo y Manolo. Sabios a los que tuve la inmensa suerte de conocer. Somos científicos huérfanos desde el día en que se fueron.

Gracias a Soraya, Santi, Nerea, Miguel, Diana, Yasmina, Héctor, Angélica, Fernando y todos los que pasaron por el "Roca" y con los que he compartido fabulosos momentos. Gracias a todos los "de arriba". Siempre hay motivos para subir esa escalera y pasar ratos inolvidables en la "Ibervending Room". Nunca tendré mejores compañeros de espectros y cerveza.

Gracias a toda la gente con la que he compartido mis estancias en el extranjero. A Rui Brito, por permitirme disfrutar de Coimbra durante cuatro meses y todo su apoyo en los últimos momentos del máster. Gran investigador y mejor persona. Gracias de todo corazón a Ann McDermott por retarme a aprender RMN de sólidos en tres meses, poniendo a mi disposición ilimitados recursos para que el tiempo no fuera una excusa. Sigues sin imaginarte cuánto te admiro y todo lo que aprendí de ti. Nueva York ya no es un simple amasijo de cemento. Gracias a todos mis compis de Columbia que hicieron de "La Jungla" un hogar a medida: Yunyao, Keith, Ivan, Rivkah y Gary.

Gracias Luci, Rui, Mai, Alberto, Juan Pablo, Ana. Empezamos como compañeros de máster y sin darnos cuenta, nos hemos convertido en compañeros de vida. Amigos, os debo un gran parte de mi sonrisa.

Gracias a todos mis amigos, que siempre están ahí pese a la distancia. Toni, Peri, Guillo, Donna, Maydo, Clara… Sois increíbles.

Gracias a toda mi familia. A Juan Carlos, por hacer de la vida compartida algo tan bello y estar siempre a mi lado. A mi hermano, por saber cuándo necesito una buena pinta de Guinness o un hombro donde llorar. ¡Qué enorme saber que estás ahí!
Y a mis padres. Por tejer serenidad y haber sabido quererme, enseñarme, motivarme. Miro atrás y veo lo bien que lo hicieron en cada momento. Soy quien soy gracias a ellos.

Vuestros nombres son faros en la noche,
vuestra voz en la distancia un ansiolítico.
Mi amor hacia vosotros, el eco del amor que recibo.

La ciencia no sirve sino para darnos una idea de cuán vasta es nuestra ignorancia.

Félecité de Lamennais

CONTENTS

LIST OF APPENDICES **197**

LIST OF ABBREVIATIONS

αS:	Alpha-synuclein
Aβ:	Beta-amyloid Peptide
AD:	Alzheimer's Disease
ALS:	Amyotrophic Lateral Sclerosic
APP:	Beta-amyloid Precursor Protein
CD:	Circular Dichroism
CFTR:	Cystic Fibrosis Transmembrane Regulator
CPEB:	Cytoplasmic Polyadenylation Element Binding
CR:	Congo Red
DFT:	Density Functional Theory
DNA:	Deoxyribonucleic Acid
FTLD:	Frontotemporal Lobar Dementia
FTLD-TDP:	Histological subtype of FTLD associated with TDP-43
FUS/TLS:	Fused in Sarcoma/Translocated in liposarcoma
HBC:	Hydrogen Bond Cooperativity
HBS:	Hydrogen Bond Strength
HD:	Huntington Disease
hnRNP:	heteronuclear Ribonucleoprotein
Hsp104:	Heat Shock Protein
LC:	Low Complexity (sequences)
MD:	Molecular Dynamics
mRNA:	messenger RNA
NMR:	Nuclear Magnetic Resonance
NOE:	Nuclear Overhauser Effect
NTD:	N-terminal Domain
PD:	Parkinson's Disease
Pmel17:	Premelanosome Protein
QM:	Quantum Mechanics
REF:	Rubber Elongation Factor
REMD:	Replica-Exchange Molecular Dynamics
RHIM:	RIP Homotypic Interaction Motif
RIP:	Receptor-Interacting Protein
RMSD:	Root-Mean-Square Deviation
RNA:	Ribonucleic Acid
RRM:	RNA-binding Motif
SG:	Stress Granule
ssNMR:	solid-state NMR
STEM:	Scanning-Transmission Electron Microscopy
TDP-43:	Transactive Response DNA Binding Protein 43 kDa
TEM:	Transmission Electron Microscopy
ThT:	Thioflavin-T
TTR:	Transthyretin
XRFD:	X-ray Fibril Diffraction

Resumen

El amiloide es un tipo de estructura de proteínas consistente en láminas beta muy largas y altamente estables, responsables de más de veinte enfermedades mortales en humanos, incluyendo Alzheimer, Parkinson y Esclerosis Lateral Amiotrófica (ELA). Paradójicamente los amiloides no siempre son patológicos, sino que también son esenciales en procesos vitales tales como la regulación del ARN o la consolidación de la memoria. A pesar de los avances en el campo, todavía hay muchos detalles desconocidos acerca de los aspectos energéticos y la formación de estas estructuras.

En esta tesis presento resultados sobre la formación y los factores que estabilizan el amiloide del heptapéptido GNNQQNY de Sup35, un prion de levaduras. Mis resultados computacionales y experimentales muestran que el intermedio clave en la amiloidogénesis es una lámina beta paralela curvada compuesta por tres hebras. Esta lámina es capaz de dimerizar expulsando fácilmente las moléculas de agua de solvatación y así producir la estructura llamada columna beta cruzada (cross-β spine). La presencia de enlaces de hidrógeno híper-estables es crucial para la estabilidad tanto de este intermedio, como de oligómeros más largos y fibras de Sup35. Con estos datos, se propone un modelo para la formación de este núcleo que plantea la cuestión de si los oligómeros tóxicos son una consecuencia de columnas beta cruzadas en un plegamiento anómalo.

La TDP-43 (Proteína de unión a ADN de respuesta transactiva de 43 kDa) es particularmente fascinante, puesto que su región C-terminal forma agregados de tipo amiloide implicados en la ELA, además de un hipotético amiloide funcional esencial para el transporte y la regulación del ARN. Puesto que existe un debate acerca de la naturaleza amorfa o amiloide de los agregados de TDP-43, en esta tesis se presentan una serie de ensayos bioquímicos y espectroscópicos que revelan la naturaleza de tipo amiloide de estos agregados. Entre ellos se incluyen experimentos de unión a cromóforos y fluoróforos, reconocimiento por anticuerpos conformacionales, dicroísmo circular, difracción de rayos-X, resonancia magnética nuclear (RMN) en estado líquido y sólido, así como microscopía electrónica.

Mediante métodos espectroscópicos y computacionales, hemos investigado conformaciones permitidas para el posible amiloide patológico, elaborando modelos atómicos para el monómero constituyente. La estructura consiste en horquillas beta alineadas en una topología de giro beta y empaquetadas en paralelo para formar una estructura cuasi-amiloide con un plegamiento novedoso.

Por otra parte, el dominio N-terminal (DNT) de la TDP-43 juega papeles clave en la regulación de la agregación funcional y patológica de la región C-terminal. La estructura de alta resolución del DNT ha sido obtenida por RMN y consiste en una hélice alfa y seis hebras beta, con un núcleo hidrófobo bien empaquetado, dos grupos de residuos cargados negativamente y dos cisteínas expuestas y distantes entre sí. La estabilidad conformacional, determinada por intercambio de hidrógeno/deuterio, es de 4 kCal/mol a pH 4 y 25 °C. Las propiedades dinámicas han sido elucidadas usando métodos de RMN y muestran que la hélice alfa y cinco de las seis hebras beta son rígidas. Una de las hebras del borde y los giros son más flexibles. Estos descubrimientos avanzan nuestra comprensión del DNT y proveen un medio para estudiar su interacción con la región C-terminal.

Abstract

Amyloid is a class of protein structures composed of very long and highly stable β-sheets. Amyloid-like oligomers cause over twenty mortal human diseases, including Alzheimer's, Parkinson's diseases and Amyotrophic Lateral Sclerosis (ALS). Paradoxically, other amyloids are key to vital physiological processes, including RNA regulation and memory consolidation. Despite advances, many details on the formation and energetics of amyloids are still unknown.

In this thesis, I present results on the formation and factors that stabilize the amyloid formed by the yeast prion Sup35 heptapeptide GNNQQNY. My computational and experimental results revealed that the key intermediate in amyloidogenesis is a twisted, three-stranded, parallel β-sheet. This sheet dimerizes with facile release of water molecules to yield the cross-β spine structure. Hyper-stable H-bonds are crucial to the stability of this intermediate and to that of larger Sup35 oligomers and amyloids. A model for this nucleus' formation is proposed, raising the question of whether noxious oligomers are a consequence of misfolded cross-β spines.

TDP-43 (Transactive response DNA-binding protein 43 kDa) is particularly fascinating because its C-terminal region forms amyloid-like aggregates implicated in ALS yet also a putative functional amyloid vital to RNA regulation and transport. Since a debate existed as to whether TDP-43 aggregates are amyloid-like or amorphous, a series of biochemical and spectroscopic assays were performed, which revealed the amyloid-like nature of these aggregates. This includes fluorophore and cromophore binding, recognition by conformational antibodies, circular dichroism, X-ray diffraction, solution and solid-state NMR, and electron microscopy.

Using spectroscopic and computational methods, the conformation of the putative pathological amyloid was investigated and structural models for the monomeric amyloidogenic intermediate and amyloid-like oligomers were determined. The structure consists of β-hairpins aligned in a "β-turn" topology and packed in parallel to form an amyloid-like structure whose topology is novel.

The N-terminal domain (NTD) of TDP-43 plays key roles in regulating the functional and pathological aggregation of the C-terminal region. The structure of the NTD has been solved at high resolution by NMR methods, and it consists of an α-helix and 6 β-strands featuring a well-packed hydrophobic core, two exposed clusters of negatively charged residues and two separated, exposed cysteine residues. The conformational stability, determined by hydrogen/deuterium exchange, is 4 kcal/mol at pH 4 and 25 °C. The dynamic properties were elucidated using NMR methods, and show that the α-helix and 5 of 6 β-strands are rigid. One edge strand and the loops are more flexible. These findings advance our understanding of the NTD and provide the means to study its interaction with the C-terminal region.

Chapter 1

Introduction

This Chapter begins with an overview on the concept of *amyloid*, a class of protein structures composed of very long and highly stable β-sheets, describing their contrasting pernicious and physiological functions in organisms ranging from bacteria to human. Next, the structural features of the GNNQQNY are described in detail as a representative example of amyloid fibrils to provide background for the results presented in Chapter 2. Some of the recently published high-resolution solid-state NMR amyloid structures are also shown to illustrate subtle differences with respect to the canonical fold, since these atomistic models are expanding our knowledge of these fibrillar aggregates.

Once these concepts are introduced, an overview of the TDP-43 protein will be given, since this RNA processing and regulating protein is the major culprit in amyotrophic lateral sclerosis (ALS) and has also been linked to frontotemporal lobar degeneration (FTLD-TDP43) and Alzheimer's disease. The overview of TDP-43's domains organization and function will provide a conceptual foundation for the research presented in Chapters 3 to 6. Finally, the last section will provide an integrative summary of the Thesis and highlight its most relevant findings.

1.1 The amyloid dichotomy: pernicious and functional aspects

1.1.1 Pernicious amyloids

Amyloids are insoluble proteins that were originally identified in certain systemic diseases, where proteinaceous aggregates build up, blocking the function of vital organs and causing amyloidosis. These amyloids can affect different organs, ranging from the heart, kidneys or liver to the digestive tract or the eyes, and although they are not toxic to the cells they contact, normal physiological functions can be sterically blocked. Transthyretin (TTR), a 55 kDa homotetramer which transports retinol and thyroxine through the serum and cerebrospinal fluid, is an exemplary amyloid. When the tetramer dissociates, monomers can misfold into an aggregation-prone conformation that results in amyloid formation and blindness. Remarkably, these aggregates are not toxic to retinal neurons as evidenced by the fact that replacement of the vitreous humor with the appropriate optically clear medium fully restores the vision [1].

In the 1980's, Glenner and others identified that the principal component of the plaques which are a classical clinical hallmark of Alzheimer's disease (AD) is an amyloid fibril composed of the peptide Aβ [2, 3]. Unlike the nontoxic amyloids associated with classical systemic amyloidosis, Aβ forms amyloid-like soluble oligomers which are highly toxic [4]. These pernicious species are in equilibrium with long, mature amyloid fibrils and can be found in a wide range of sizes and topologies (see **Fig. 1**). Since this finding, a score of other proteins capable of forming this second type of pernicious amyloids have been identified in the context of human pathogenesis [5] Many are of special relevance in neurological disorders [6] For example α–synuclein, huntingtin, and TDP-43, are involved in Parkinson and Huntington diseases, and Amyotrophic Lateral Sclerosis (ALS), respectively. Unlike systemic amyloidosis, these neuronal disorders cannot be treated by removing amyloid deposits or organ transplants. Therefore, improving our knowledge on this field is essential.

It is important to note that, from a clinical point of view, the term *amyloid* is restricted to the intrinsically, nontoxic accumulation of systemic deposits and is not recommended for these aggregation-prone proteins that form toxic intermediates [1]. However, from a biophysical perspective, the term can be safely applied to both types of protein aggregates [6] and in this Thesis, I focus on the second category of amyloids in which pernicious intermediates usually appear during fibril formation.

How the same quaternary structure forms from different proteins involved in different processes remain elusive, and the more recent discoveries that certain amyloids are not only nontoxic but perform vital physiological functions has opened new challenges for our understanding of this paradigmatic fold [7, 8].

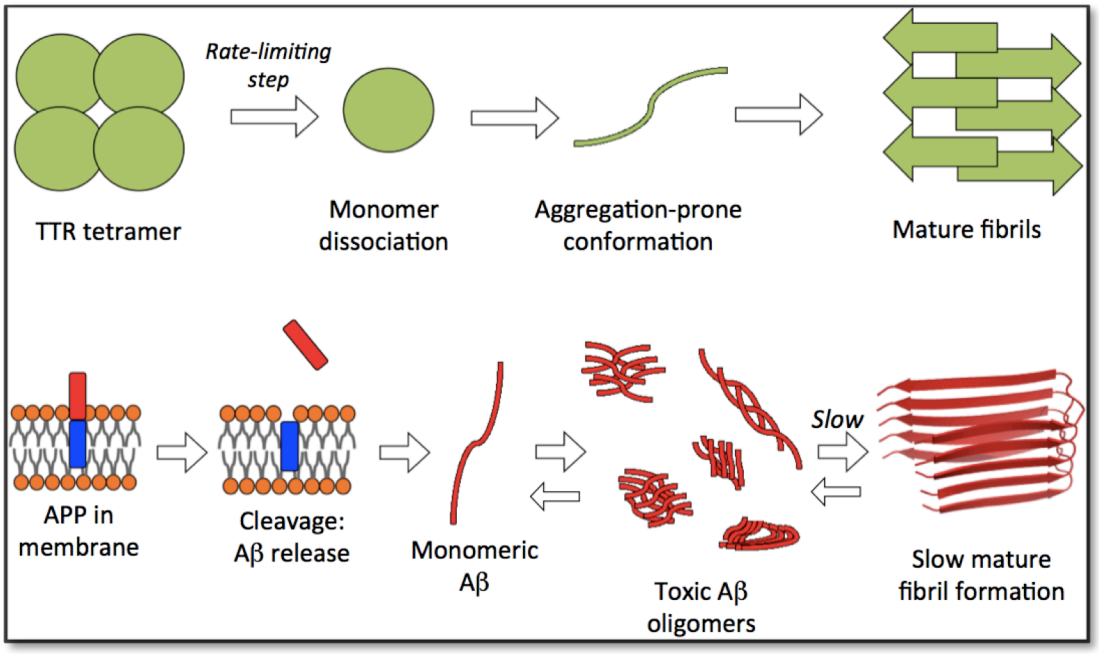


Fig. 1. **Models for toxic oligomers and nontoxic amyloid formation.** *Top*: TTR amyloid formation requires monomer dissociation from the native tetramer that evolves into an aggregation-prone conformation that allows fibril formation. These fibers are nontoxic themselves, but their progressive accumulation affects organ's vital function. *Bottom*: APP is cleaved to yield Aβ fragments that rapidly assemble into toxic oligomers of different sizes and topologies. Fibril formation is a slow process, which yields less toxic amyloids.

Aβ oligomers trigger Alzheimer Disease – Alzheimer disease is a progressive neurodegenerative illness that causes loss of neurons and synapses in the cerebral cortex, with harmful amyloid plaques and neurofibrillary tangles formation in the brain.

In early stages of the disease short-term memory loss occurs, which is followed by disorientation, problems with language and loss of motivation, among other behavioral aspects. It gradually worsens until death, generally eight years after their symptoms are noticeable.

The link between the Aβ peptides and AD was uncovered when Glenner identified this polypeptide as the major component in brain inclusions of patients suffering the disease [2]. Moreover, pathological mutations located in the Aβ parent protein or within Aβ lead to increased amounts or oligomerization and early disease onset. Another observation was the mapping of the Aβ gene to chromosome 21 [9], which explained why people with trisomy of chromosome 21 (Down's syndrome), having three "doses" of this gene, develop Alzheimer's disease in their thirties.

Aβ peptides are generated from the β-amyloid precursor protein (APP), which is sequentially cleaved by two secretases to yield ~4 kDa fragments that are the major components identified in amyloid deposits in the brain of patients with AD. [10]. Among the different polypeptides produced, Aβ40 is the more abundant (80-90 % of all species) and less aggregation prone, whereas Aβ42 is a minor (5-10 %), more hydrophobic fragment (**Fig. 2**) considered to be the most pathogenic among the Aβ

species found in AD [11, 12]. Why these fragment are generated is still poorly understood, but eliminating Aβ all together does not seem to be an appropriate strategy to prevent AD development, as there is evidence that it is essential for the development of the myelin sheath [13] for certain peptide-receptor interactions [14] or even to have a potent bactericidal activity, which might be its true physiological functions [15]. Recent evidence suggests that Aβ toxic aggregates can convert disordered Aβ into pathological conformation and spread perniciously in a prion-like manner via axons [16].

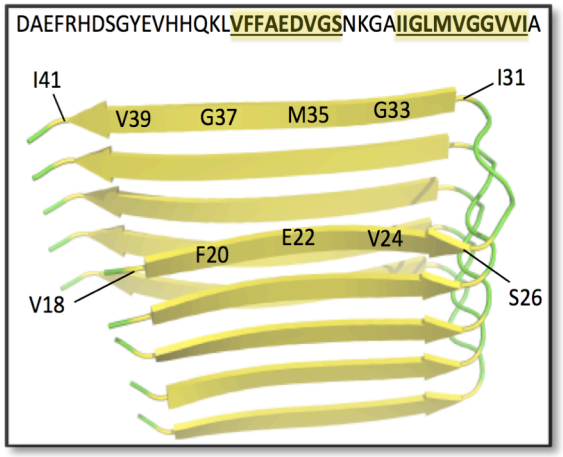


Fig. 2. Structural model for Aβ(1-42) amyloid fibrils. The Aβ peptide is predominantly hydrophobic as can be seen in its sequence. Residues underscored and highlighted in yellow show protection against H/D exchange [17] and are involved in β-sheet formation. This structural model lacks the first sixteen residues to illustrate the β-strands of the amyloid core. This structure is deposited in the PDB under accession code 2BEG.

Parkinson Disease is a consequence of α-synuclein misfolding – Parkinson's disease (PD) is a disorder of the central nervous system that impairs dopamine production in the substantia nigra of the brain. As a result, progressive lost of movement and coordination starts, featuring symptoms such as shaking, slowness of movement, muscular rigidity and postural instability. The pathology of the disorder involves formation of inclusions known as Lewy bodies in neurons that, in parallel with AD, are predominantly composed of an aggregated, amyloid form of a protein (α-synuclein, αS). This 140-residue protein adopts a helical conformation upon phospholipid membrane binding [18], a process which is carried out by the presence of lipid-binding motifs in its N-terminal domain (residues 1-60). The C-terminal region is enriched in negatively charged residues, which might be responsible of the solubility and the unstructured properties of the protein in solution. α-Synuclein has a middle domain composed of residues 61 to 95, which is absent in the other members of this family of proteins (β- and γ-synuclein), and is responsible of its oligomerization into β-rich aggregates. This region, called the *non-Aβ component,* can form amyloid fibrils under incubation and is therefore directly linked to PD development [19].

It is remarkably that both the two different polypeptides linked to AD and PD form highly toxic intermediate oligomeric species during amiloidogenesis, which are more toxic than mature fibrils [20–22]. The structural similarity of these oligomers from Aβ and α-synuclein is underscored by their recognition by a common conformational antibody [23], as well as their exceptionally high mechanical stability [24]. Despite these common features, they are different diseases caused by different proteins and therefore, there are a number of differences between both pathologies at the Cell Biology level [12]. Nevertheless, the remarkably similarities strongly suggests that improving our basic understanding of amyloid formation and stability could lend insight into the origin of multiple diseases.

Amyloid formation in Huntington disease is 100% genetically encoded – Huntington disease (HD) is a rare neurodegenerative disorder that causes involuntary movements known as chorea. It is also characterized by a loss of cognition, along with changes in personality and psychiatric manifestations with a fatal dénouement, with life expectancy of 15 to 20 years after symptoms begin. Like AD and PD, Huntington disease is characterized by the accumulation of amyloid fibrils in the brain. However, whereas most Alzheimer and Parkinson cases are stochastic and have not genetic component, all cases of Huntington's disease are due to an abnormal large number of glutamine residues within the Huntingtin protein (**Fig. 3**). [25, 26].

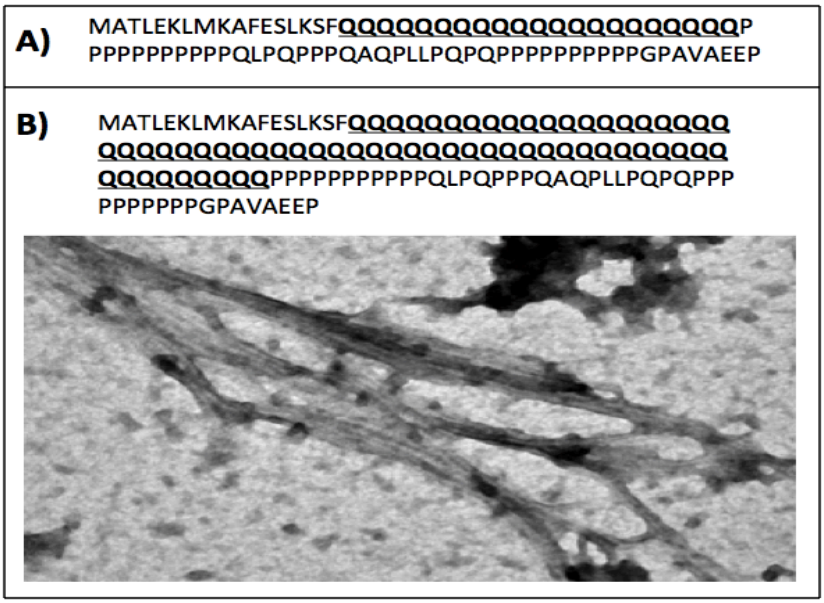


Fig. 3. **Polyglutamine repeats trigger Huntington disease.** The first 60 residues of the Huntingtin protein are shown in (A), displaying 21 sequential glutamines. Panel B shows a pathological scenario where the number of Gln residues has increased to 62. As a result, amyloid fibril formation occurs, as shown by electron microscopy. This Figure is taken from Chapter 4.

Huntingtin is a large protein (347 kDa) whose precise function in humans remains elusive, although it is known to be involved in several processes such as microtubule-mediated vesicle transport [27]. Its N-terminal domain (NTD) has a repetitive segment with 11 to 34 glutamine residues. Slippage of the polymerase during DNA replication can lead to an increase in the number of these residues, and 35

consecutive Gln's is considered to be the threshold for developing the disease. Beyond this limit, huntingtin aggregation into amyloid fibrils is triggered with fatal consequences, and the higher the number of consecutive Gln's, the earlier the onset of the disease. It is interesting that polyglutamine repeats are also known to cause other rare illnesses [28], and this might not be a particularity of Gln residues, since polyalanine stretches has been found in other diseases[29]. Although they are not discussed in this Thesis, it is noteworthy that the toxic threshold for polyAla's repeats is smaller than that in polyGln's. From these observations, it is clear that misfolding and aggregation of otherwise soluble proteins into amyloid-like oligomers and fibrils cause Alzheimer's, Parkinson and Huntington diseases.

TDP-43 aggregates are involved in ALS and FTLD – TDP-43 is a heteronuclear ribonucleoprotein (hnRNP) that has attracted much interest since it was shown to be the major component of the hyperphosphorylated, polyubiquitinated proteinaceous aggregates in patients with ALS and FTLD [30]. In ALS, death of motor neurons impairs the ability of the brain to control muscle movement and as a consequence, people gradually lose the ability to speak, move, eat, and finally, to breathe. This disease is the third most common neurodegenerative disorder after AD and PD.

FTLD is the result of a progressive neuronal loss in the frontal and/or temporal lobes, and is difficult to differentiate from Alzheimer's disease in the early stages since their initial development is very similar. When accumulation of TDP-43 aggregates is the histological feature of the disease, it is referred to as FTLD-TDP since there are different subtypes linked to the major presence of other proteins.

TDP-43 is widely studied during this thesis and therefore, it deserves a special discussion. In this Chapter, *Section 1.3* is dedicated to this protein's structural domain organization and functions. However, it is now advanced that aggregates formed by full-length TDP-43 have been recently shown to be composed of amyloid fibrils [31], and this finding shed light on a large debate about the nature of this protein's inclusions. The presence of a prion-like segment within the C-terminal region of TDP-43 and its structural characterization is discussed later on, as it might be the kernel for pathological aggregates formation in ALS and FTLD-TDP.

1.1.2 Functional amyloids

Regardless of their implication in pathological scenarios, amyloids are not only involved in detrimental processes, but they also play beneficial roles in many organisms. These fibrillar species are the so-called *functional amyloids,* and different proteins in a wide range of organisms can form these structures. For instance, in *E. coli* or *Salmonella* spp., curling proteins assemble into amyloid fibrils that form the fimbriae or pilli, an appendage that allow cell and surface adherence. The presence of these amyloids is used for surface colonization, the main mechanism in host invasion. This oligomerization is regulated by at least six proteins, suggesting that this process is highly controlled and regulated [32–34]. This highlights the complexity of amyloid formation even in organism as simple as bacteria, and seems to contradict the idea of stochastic misfolding and aggregation.

Amyloids from yeasts are known to represent a mechanism for phenotype diversity generation. In higher-order organisms, including humans, proteins that form functional amyloids adopt this fold to carry out different essential roles that range from UV protection to consolidation of memories. This raises the possibility that the pernicious amyloids previously described my result from misregulation of functional amyloid formation. On the basis of this hypothesis, one could speculate on the dichotomy of Aβ having bactericidal role and causing Alzheimer's disease, for instance. Some hypotheses in this vein are sketched out at the end of this section. However, much research is still needed for withdrawal of solid conclusions, even though a number of functional amyloids is already reported (see **Table 1**) and surely more and more examples shall be advanced.

Sup35 amyloid ensures genetic variability in yeasts – The most thoroughly studied functional amyloids are those formed by yeast proteins. One of the best understood is Sup35, a 685-residue protein. The large C-terminal region of Sup35 is implicated in translation termination and ensures that protein synthesis ends at stop codons. The middle domain of this protein is enriched in charged residues, which promote Sup35 solubility. Interestingly, the N-terminal domain possesses a high content of Asn and Gln residues that may be responsible for the transition from of the soluble species into an aggregated, amyloid form. This fibrillar state prevents translation termination and generates phenotypic diversity, most likely due to the production of an altered, complex proteome [35]. The finding that a heat shock protein (Hsp104) is able to break these fibers into smaller fragments, which seemt to act as seeds to guarantee inheritance of this phenotype by daughter cells, suggest that this process is controlled to be triggered under specific conditions [36].

Although initially controversial, the identification of prions as genetic elements posing non-Mendelian inheritance to confer phenotypic feature to wild yeasts is now gaining acceptance [37].

Fungal prions dictate strain survival –HET-s protein from the fungus *Podospora anserina* [38] is another interesting functional amyloid. HET-s contains 289 residues organized into a well-defined, compact domain (residues 1-227) and a C-terminal aggregation-prone region (residues 218-289) [39] that form amyloid fibers, as revealed by the high-resolution structures obtained by solid-state NMR [40, 41]. In fact, deletion of this segment prevents amyloid formation *in vitro* and propagation *in vivo* [42]. HET-s' transition from a disordered conformation into a functional amyloid mediates a type of cell death reaction, called heterokaryon incompatibility [43, 44], which results in the inability of two strains to fuse unless the alleles of this gene are identical. To carry out this function, two alleles exist in the het-s locus that encode the HET-s (infectious form) and HET-S (soluble form) proteins, which differ in 13 amino acids [45]. When a [Het-s] cell fuses with a [Het-S] cell, the resulting diploid cell dies (**Fig. 4**). It is thought that the aggregated state impedes proteins from binding to their targets, resulting in a phenotypic loss of function.

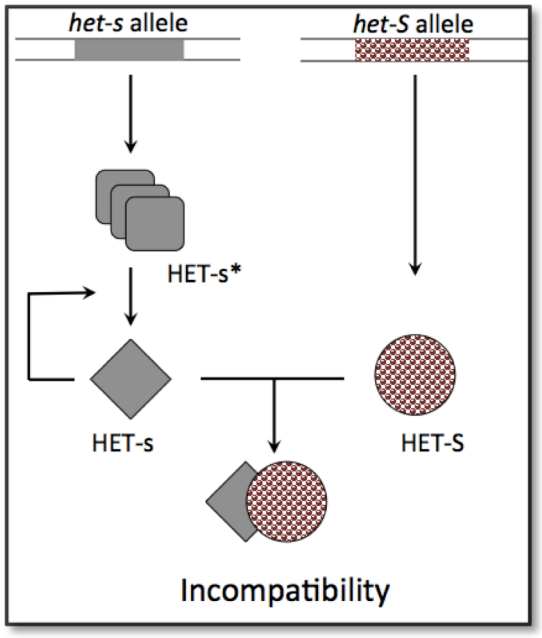


Fig. 4. Model for heterokaryon incompatibility mediated by functional amyloid formation. The *het-S* allele is responsible of the soluble form of HET-S protein, which is not involved in heterokaryon incompatibility. However, the HET-s protein, encoded in the *het-s* allele, can undergo a structural transition and form a functional amyloid. Initially, this protein is not directly involved in cell death (HET-s*). However, it can convert into an infectious state (HET-s) that recruits all HET-s* into this conformation and triggers the reaction against HET-S to cause heterokaryon incompatibility. Adapted from reference [38].

Functional amyloids in higher-order organisms – Functional amyloids are even found in plants, as in the case of the Rubber Elongation Factor (REF), a 138-residue protein that mediates the biosynthesis of natural rubber. This REF protein has been shown to form fibrillar aggregates [46], although its role in the amyloid state has not been fully elucidated. Considering its implication on the isoprenoid biosynthetic pathway, it could be speculated that this amyloid may protect cells by inhibiting the release of toxic metabolic intermediates. Interestingly, there is a well-characterized functional amyloid in humans that also seems to perform a protective function. This protein, Pmel17, is involved in melanin organization in mammalian cells, [47] which is described in the next paragraph.

Pmel17 forms a functional amyloid responsible of UV-protection in mammals – Pmel17 is a 661-residue transmembrane glycoprotein, which is processed inside membrane-limited organelles (melanosomes) to release an 80-kDa [48], self-assembly segment that form amyloid fibrils and provides a scaffold for melanin synthesis [49, 50]. To reduce the danger of toxic amyloid-like oligomers, Pmel17 oligomerization only begins following a specific proteolytic cleavage and occurs only at low pH within a specialized vesicle, the melanosome. Moreover, the rate of fibrillization within the melanosome is four order of magnitude faster than that of Aβ or α-synuclein, which might prevent the build up and diffusion of harmful metastable amyloid oligomers. Besides organizing melanin, the Pmel17 amyloid network might also prevent the release of toxic intermediates of melanin biosynthesis; namely, indole-5,6-quinone (DHQ). Interestingly, these molecules resemble the amyloid specific fluorophore Thioflavin-T, which suggests a structural link between toxic and physiological amyloids.

A human amyloid acts as a signaling complex to activate programmed necrosis – The RIP1 and RIP3 proteins (Receptor-Interacting Proteins 1 and 3) are a group of serine/threonine kinases that mediate death-inducing processes in cells [51, 52]. These proteins have RIP homotypic interaction motifs (RHIMs) located in the intermediate domain of RIP1 and in the C-terminal tail of RIP3 that are responsible for the interaction between these two kinases [53]. These regions, which are rich in Asn, and Gln, as well as hydrophobic residues, mediate the formation of a functional amyloid, as shown in **Fig. 5**. This complex is essential for programmed necrosis to occur, as have been shown *in vitro* and by isolating the RIP1/RIP3 complex from necrotic cells [54].

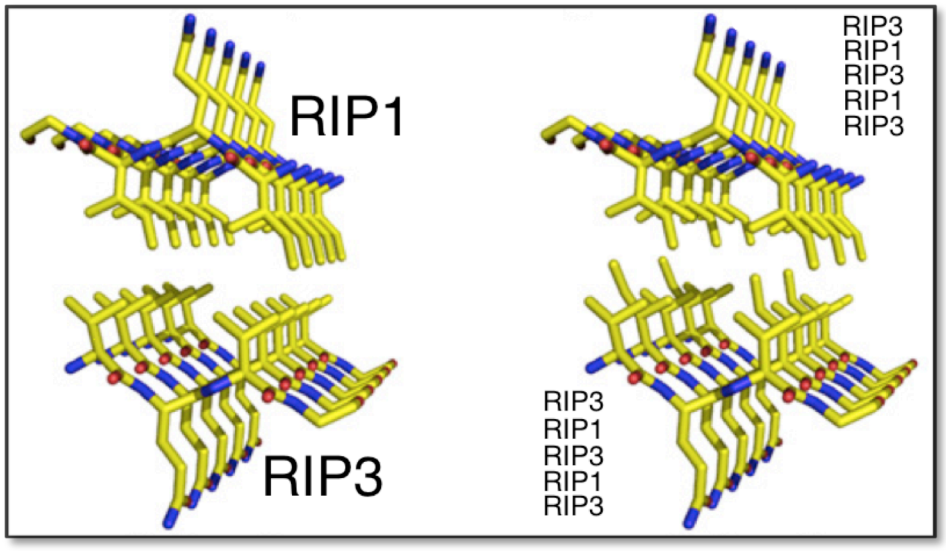


Fig. 5. **Structural models proposed for the RIP1/RIP3 functional amyloid.** The assembly of these to proteins to form a signaling functional amyloid and trigger necrosis in cells is mediated by the formation of a hydrophobic core. To models have been proposed: (*left*) two mating β-sheets, each composed of the same protein, or (*right*) heterogeneous sheets formed by the stack of alternating RIP1 and RIP3 molecules. Reproduced from reference [54].

Functional amyloids consolidate long-term memories – A neuronal isoform of the Cytoplasmic Polyadenylation Element Binding (CPEB) protein was found in the sea snail *Apysia californica* to form a functional amyloid involved in memory consolidation [55]. Apart from the RNA binding domains, this isoform contains a prion-like N-terminal domain with a remarkably high content of Asn and Gln residues, that is key for long-term memory consolidation [56, 57]. In *Apysia*, CPEB modulates its function in protein synthesis at synapse through a transition from a soluble α-helix-rich oligomer into stable prion-like state with enhanced affinity to bind target mRNAs [58] In *Drosophila* the homolog is called Orb2 [59], and it is present in two isoforms: Orb2A and Orb2B. Despite its low abundance and instability, Orb2A is capable of self-assembling into functional amyloid fibrils essential for memory persistence [60]. Interestingly, CPEB proteins also exist in humans and other mammals and have been recently shown to form functional amyloids key for memory consolidation [61]. **Fig. 6** illustrates a model for functional amyloid formation upon neurotransmitter stimulation.

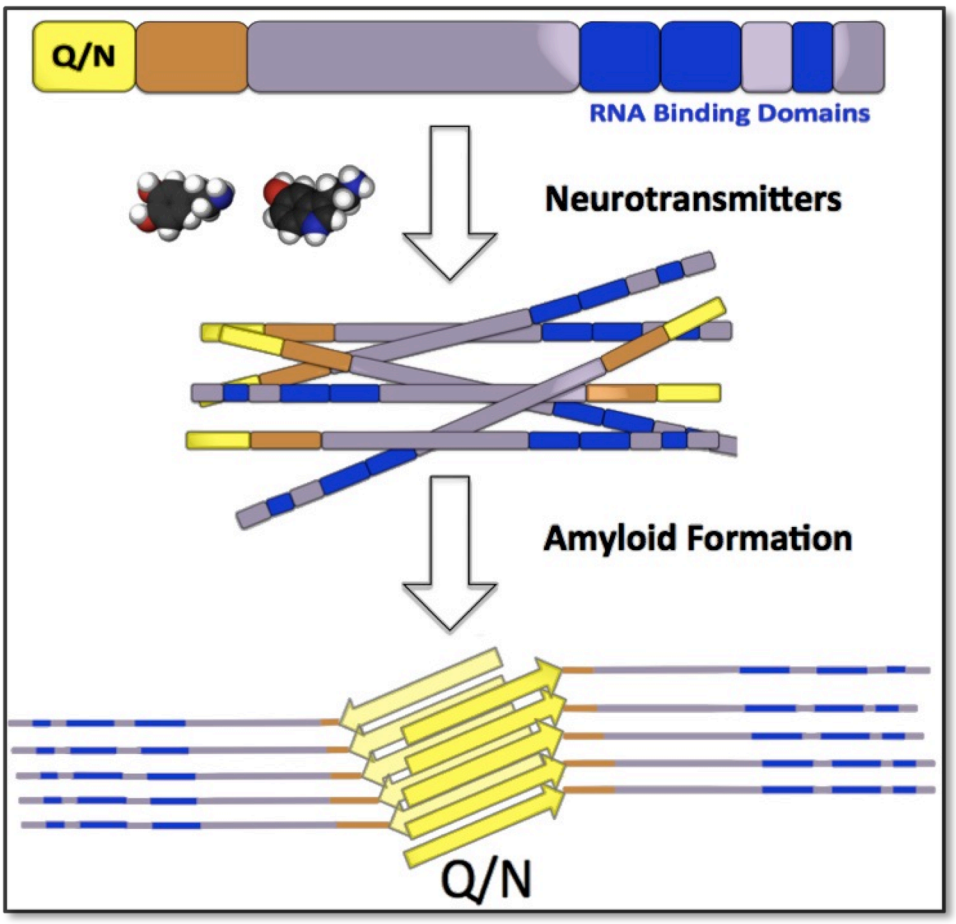


Fig. 6. Amyloid Formation in CPEB. Upon neurotransmitters-mediated stimulation, an increased expression of CPEB occurs followed by amyloid formation through the prion-like domains of CPEB. This functional assembly has enhanced mRNA binding affinities and changes RNA regulation, triggering enduring structural changes in the synapse which are related to long term memory consolidation.

FUS: possible role in stress granule formation – Like CPEB, Fused in Sarcoma/Translocated in Liposarcoma (FUS/TLS) is an RNA-binding functional amyloid. This 526-residue heteronuclear ribonucleoprotein (hnRNP) performs a wide range of functions, including processing and transport of mRNA out of the nucleus. FUS is composed of several distinct domains, including one S,Y,G,Q-rich region (SYGQ), an RNA recognition motif (RRM), three R,G,G-rich regions (RGG1, RGG2, RGG3) and Zinc finger motif (ZnF) [62]. The C-terminal sequence functions as a nuclear localization sequence [63], and mutations located in this region result in mislocalization of the protein in the cytoplasm, which can result in harmful aggregation [64]. In contrast, where the substitutions are located in the N-terminal SYGQ region or in the RGG regions, the nuclear localization of FUS is not altered, at least in cultured cells [65]. Some FUS mutations have been related to familial ALS cases [66]. Especially, G156E has been shown to form amyloid fibrils and triggered aggregation of full-length FUS [67].

Stress granules (SGs) are dense cytosolic aggregates composed of proteins and RNAs that are formed when the cell is under stress. The formation of these membrane-free inclusions serves to rapidly and reversibly prevent mRNA translation under such conditions. They can also act transporting mRNA to concrete regions where local protein synthesis is required. FUS/TLS has been recently proposed to be involved in this transporting mechanism through the formation of "hydrogels" composed of amyloid fibrils [68]. This is achieved through [G/S]-Y-[G/S] tripeptide repetitions called low-complexity (LC) sequences, that can reversibly assemble and disassemble into functional amyloids in a phosphorylation-dependent process (**Fig. 7**).

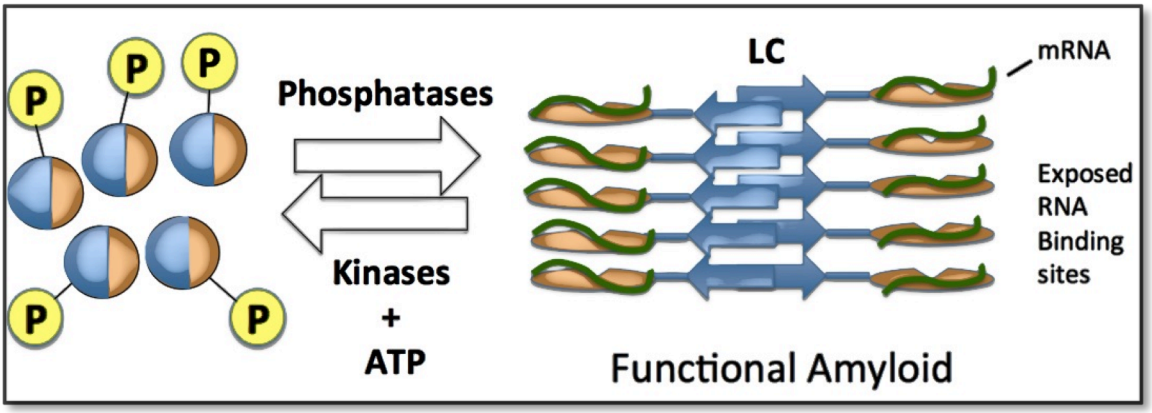


Fig. 7. **Model for amyloid-mediated stress granule formation.** Hydrogel formation is achieved by the LC sequences present in FUS. This aggregation into functional amyloids, regulated by phosphorylation, might allow extensive exposure of the RNA-binding sites, recruiting many RNA molecules and providing the scaffold for stress granule formation, which allows RNA molecules to be stored under taxing conditions or transported within these granules to specific locations. The reversibility of this process would ensure the facile release of RNA upon amyloid dissociation.

TDP-43 might form functional amyloids – The [G/S]-Y-[G/S] LC sequences in FUS might not be unique. There is a yeast ortholog of the mammalian Nup62 nuclear pore glycoprotein, called Nsp1p, which was shown some years ago to able to form hydrogels [69, 70]. This protein contains a number of FG repetitions that are essential for the assembly. On the basis of these observations, it can be speculated that TDP-43, whose important roles in ALS and FTLD were mentioned above, might also be involved in the formation of a functional amyloid, since the Q/N prion domain is flanked by [S/G]-[Ar] regions, where "Ar" denotes an aromatic residue. The sequences for these three proteins are shown in **Fig. 8**. It is intriguing that both TDP-43 and FUS/TLS are members of the heteronuclear ribonucleoprotein (hnRNP) family whose pathological aggregates are implicated in ALS and FTLD. Whether these segments, which putatively form functional amyloids, might be involved in pathological amyloids is still an open question.

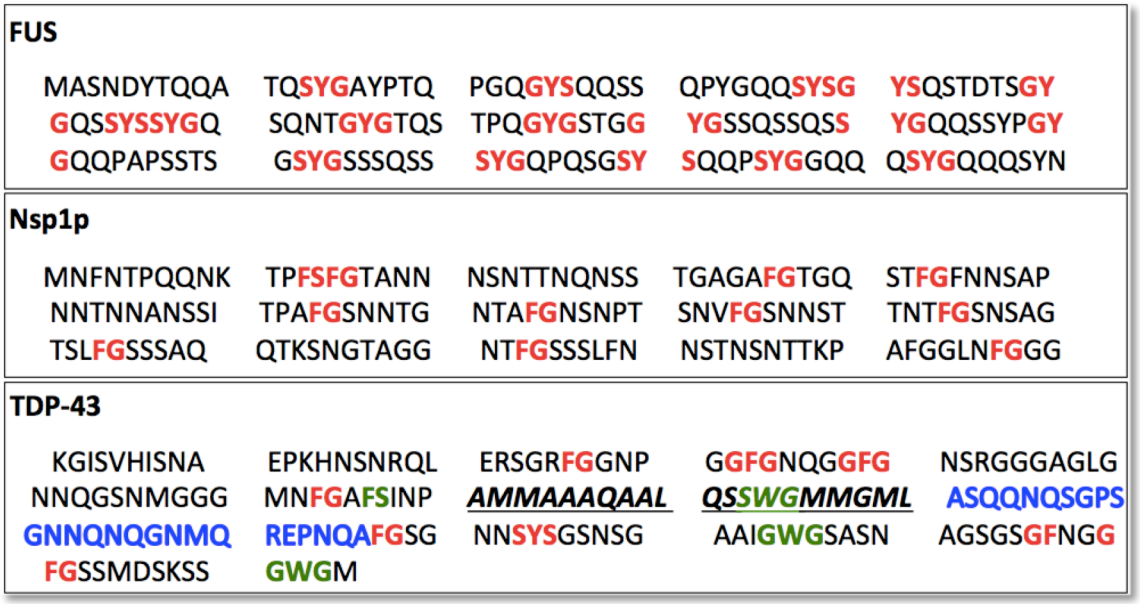


Fig. 8. Sequences of FUS, Nsp1p and possibly TDP-43, with ability to form hydrogels. Low complexity sequences reversibly form hydrogels composed of amyloid fibers, and they are proposed to be implicated in stress granule formation and RNA storage and transportation in FUS (S/G-Y-S/G repeats) and Nsp1p (FG repeats), and are colored in red. The Q/N rich region of TDP-43 (blue) is preceded by a hydrophobic patch (residues 318-330, underscored and in italics), and flanked by similar low-complexity sequences, which might suggest possible functional amyloid formation within this region. These regions are shown in red when they resemble those of FUS and Nsp1p. Note that the first FG repeat in Nsp1p is preceded by FS, also highlighted in red. While FG is the reported low complexity sequence in this protein, FS seems to be relevant on the basis of that reported in FUS. Accordingly, the color code in TDP-43 sequence is as follows: red for FUS- and Nsp1p-like (G)FG and SYS repeats, and green for similar fragments such as FS, S(G)/W/G. Tentatively, the specificity of residues comprising LC could be the fingerprint of each functional amyloids within RNA binding proteins.

Despite the overall agreement about the implication of pernicious amyloids in pathologies, the emerging examples of functional amyloids ranging from bacteria to humans (See **Table 1**) have opened new frontiers. It is fascinating that the same fold is present in all kingdoms of life yet is encoded by a vast diversity of protein sequences. In the case of Pmel17, fibril formation functions as a preventive mechanism to avoid toxic intermediates release, and provides a scaffold for melanin synthesis. On the other hand, the RIP1/RIP3 human amyloid forms a signaling complex. Thus, two very different tasks exerted by the same functional fold. It is even more astonishing that the same quaternary structure which triggers neuron death in Alzheimer's disease is essential for long-term memory consolidation.

The widespread nature of functional amyloids, the diversity of their sequences and the robustness of their fold, led Chernoff to propose that amyloid has been present and abundant since the dawn of life [79], where it could have interacted with RNA. CPEB, together with TDP-43, FUS/TLS and Sup35, strongly suggest that functional amyloid formation constitutes a "new" mechanism for RNA regulation, but is it possible that amyloid and RNA have been interacting since the epoch of the RNA World? Could functional amyloids have evolved first, and might pathological amyloids have come later, to take advantage of a successful system? Could pathological amyloids, such as prions, be the "selfish genes" of the protein world? Until pathological amyloids put in peril the future of our societies and our own personal future, we can ponder these puzzling questions about their past.

Protein (organism)	Function	Reference
Curli (*E. coli & Salmonella* spp.) TasA (*Bacillus subtilis*) FapC (*Pseudomonas* spp.) Modulin (*S. aureus*)	Biofilm formation	32–34
Mtp (M. tuberculosis)	Pili formation	71
HpaG (*Xanthomonas* spp.)	Plant pathogen virulence factor	33
Microcin E492 (*K. pneumonia*)	Storage and regulation of toxicity	33, 72
Chaplin (*S. coelicolor*)	Modulation of water surface tension	33
HET-s (*Podospora anserina*)	heterokaryon formation	38
Ure2p (*S. cerevisiae*)	Regulation of nitrogen catabolism	73
Sup35p (*S. cerevisiae*)	stop-codon readthrough	73
Nsp1p (*S. cerevisiae*)	Control of nucleocytoplasmic mixing at nuclear pore	33
REF (*Hevea brasiliensis*)	Biosynthesis of natural rubber	46
Chorion (insects and fish) pidroin (*Nephila clavipes)* Silk protein (*C. flava*)	Structural	33,74, 75
CPEB (*A. californica*)	Memory storage	56
CRES and cst8 (*Mus musculus*)	Sperm maturation	76
Pmel17 (*H. sapiens*)	Melanin synthesis	49, 50
RIP1/RIP3 (*H. sapiens*)	Regulation of necroptosis	54
Peptide hormone (*H. sapiens*)	Storage and controlled release	77

Table 1. **Functional amyloids found in different species and their functions.** Adapted from ref. [78].

1.2 The amyloid fold

All the proteins described in the previous sections form fibrils with a very similar quaternary structure, despite their diverse primary sequences. Although subtle differences exist from system to system, amyloid fibrils display a basic motif that consists of the tight pack of two long β-sheets. This is the so-called the *cross-β spine*.

The β-sheets that make up these fibrils achieve a remarkable stability due to highly polarized hydrogen bonds between the individual β-strands. The packing of two sheets is mediated by van der Waals interactions. To describe in close detail the structural features of such assembly, the X-ray structure of the GNNQQNY amyloid from the yeast prion protein Sup35 (pdb:1yjp) [80] has been chosen as a paradigm. The reason for this choice is based on its novelty, as GNNQQNY was the first amyloid structure solved at atomic resolution. Chapter 2 contains a summary of the main findings in previous works that studied GNNQQNY at atomic resolution (Table S1).

The cross-β spine fold – The crystal structure of the GNNQQNY amyloid was a milestone for studying amyloids, as it revealed their conformation at atomic resolution. The individual β-strands are bound in the β-sheets through hyper-polarized hydrogen bonds, which make them remarkably stable. It is interesting that despite this peptide's high content in asparagines and glutamines, the β-sheets bury polar residues and pack with the same remarkable efficiency as hydrophobic residues do in globular proteins. Sidechains in both sheets intermesh to achieve this packing, resembling the teeth of a zipper (**Fig. 9**). Given that this buried contact area is devoid of water molecules, the cross-β spine fold is sometimes referred to as the *dry steric zipper*. During the course of this Thesis the nature of both dry interface formation and hyperpolarization of H-bonds have been investigated, and our findings are described in Chapters 2, 4, and 6.

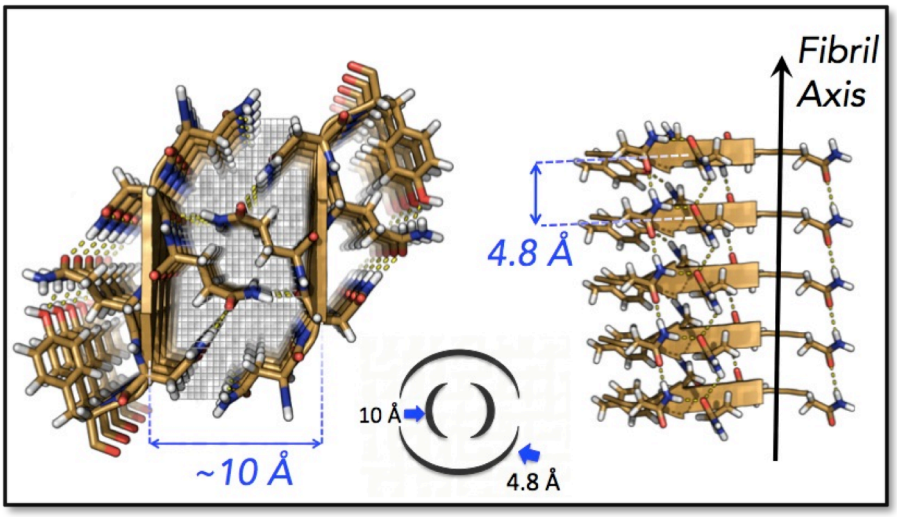


Fig. 9. Structure of the GNNQQNY amyloid from the yeast prion protein Sup35. Inter-sheet and inter-strand distances are shown at 10 and 4.8 Å, respectively. An illustration of these distances upon X-ray fiber diffraction is shown.

This particular quaternary structure constitutes a fingerprint for amyloid identification, and upon X-ray fiber diffraction (XRFD) amyloid fibrils provide a characteristic structural hallmark. This is the so-called cross-diffraction pattern, and is illustrated in **Fig 9**. Two perpendicular reflections appear at distances that correspond to the inter-strand within one sheet (~4.6 – 4.8 Å), and inter-sheet separations across the dry interface (~10 Å). The former distance has a restricted, as it corresponds to the Cα-Cα separation between two β-strands naturally constrained by their mutual backbone hydrogen bonds. Note that it is parallel to the fibril axis. The inter-sheet distance is perpendicular to the fibril axis and depends on the nature of the side chains, as their polarity and length determine the optimal interactions for the two β-sheets to appropriately pack. In Chapter 4, these distances in different amyloids are reviewed (Table S4). The 4.8 and 10 Å X-ray fibril diffraction reflections are very characteristic of amyloids; however, it is not 100% diagnostic as unusually cases of helical fibers displaying similar diffraction patterns have been reported [81].

While all amyloids feature these packing and H-bond interactions, those rich in Gln and Asn residues are able to efficiently stack and form additional sidechain-sidechain H-bonds outside the plane defined by the β-sheets. Moreover, aromatic residues (if present) also play a key role, as they can stack together and provide additional stability. These other factors exist in GNNQQNY, as shown in **Fig. 9.**

Beware of short amyloid segments – In the GNNQQNY amyloid, all the β-strands are oriented in parallel with respect to each other within the same β-sheet. Although this configuration is shared by other structures, the cross-β spine fold is not restricted to this topology, and all possible orientations of the individual monomers are allowed in principle. Accordingly, eight different classes of steric zippers have been proposed to feasibly form on the basis of geometric permutations [82], as illustrated in **Fig. 10**.

It is important to note that this classification is based on fibrils from short, amyloid-forming peptides resulting from pruning down the corresponding full-length proteins. Amyloid fibers do not easily yield well-diffracting crystals, and crystallization in the context of full-length proteins has not been achieved yet. Therefore, whereas this approach has been very useful in elucidating the essential structural aspects of amyloid fibrils, it might yield inaccurate information on the segment's conformation in a physiological context. For instance, sequences where two strands might be present within the same segment can be visualized as *beta1-loop-beta2* regions. Pruning down this segment to either β-strand1 or β-strand2 is often a productive way to obtain crystal structures, but might yield an artificial conformation. Even mixtures of both the two sheets would not necessarily give rise to the natural fold, and the situation can be further complicated considering that different crystallization conditions can perturb the configuration of the individual strands within the cross-β spines.

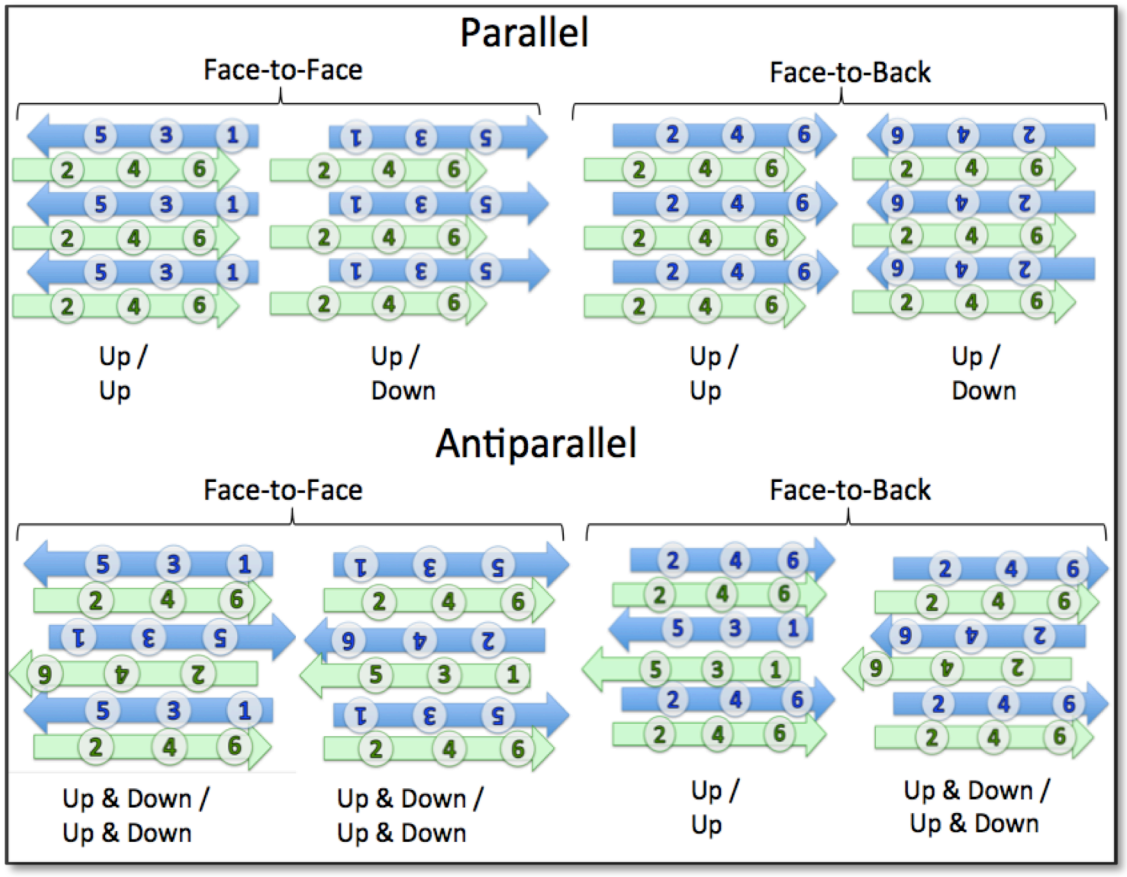


Fig. 10. **The eight possible arrangements for cross-β spine formation.** The different orientations of the peptides within the amyloid fold give rise to 2^3 possible arrangements, which can be first categorized in two groups depending on the parallel or antiparallel orientation of the strands within the mating sheets. The second division arises from the packing of the two β-sheets, depending on whether the same residues form the dry interface (face-to-face) or they are oriented in the same direction (face-to-back). This divides each of the previous groups (parallel and antiparallel) in two additional groups (parallel & Face-to-Face, parallel & Face-to-Back, antiparallel & Face-to-Face, and antiparallel and Face-to-Back). Finally, an 180º rotation of the individual strands to account for parallel or antiparallel pack of the β-sheets gives the eight possibilities shown in the figure. *Down* is used to denote the rotation with respect to the initial conformation (*Up*). A dash symbol (/) is used to separate orientations both sheets; *i.e*, sheet1 / sheet2.

The β–arc and β–turn topologies – The presence of loops connecting β-strands leads to two new classes of b-hairpins with distinct topologies, termed *β-arc* and *β-turn*. This classification is compatible with the eight possible cross-β spines of Eisenberg [82]. In the *β-arc* configuration, each monomer contributes one β-strand to the β-sheet, whereas the *β-turn* occurs when two β-strands belong to the same β-sheet. Note that to adopt the β-arc configuration, the H-bonds within each β-hairpin must break and reform with strands from a different molecule. This is illustrated in **Fig. 11**.

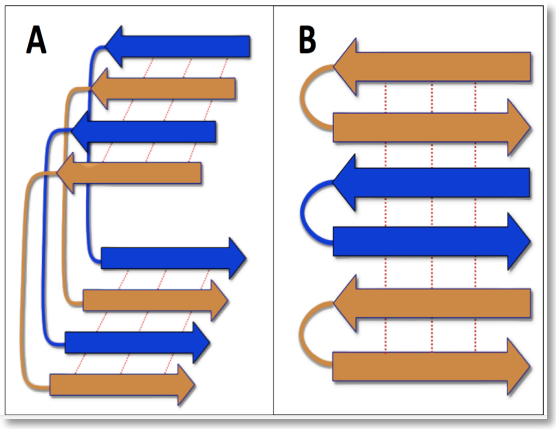


Fig. 11. β-turn and β-arc configurations. The β-turn configuration involves considering a β-hairpin as a short β-sheet formed by the two strands. Two β-hairpins in this configuration can form a dry interface (up left) or a longer β-sheet (up right). In the case of the β-arc configuration, each β-strand contributes to a different β-sheet, which requires a longer loop to pack to sheets against each other and form the dry interface.

This classification of either a β-hairpin or β-loop-β segments as β-turn/arc refers to the repetitive monomeric unit within the amyloid. An additional level of structure arises from the orientation of the loops in the amyloid-like fibril, as they can be in parallel (stacked) or antiparallel. This is illustrated in the next **Fig. 12** considering the models for Polyglutamine repeats [83] and TDP-43(341-357) [84], displaying an antiparallel and a parallel arrangement, respectively.

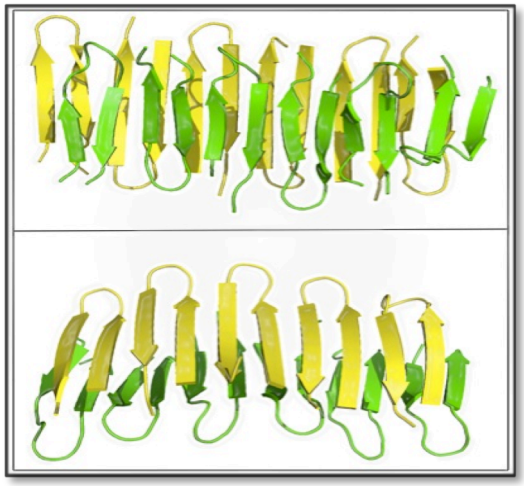


Fig. 12. Models for PolyQs and TDP-43(341-357). ***Top*****:** two β-sheets are packed against each other (green and yellow). Within each individual sheet, the loops are oriented in the opposite directions. Polyglutamine amyloid fibrils have been proposed to adopt this orientation [83].
Bottom**:** The same two β-sheets are now arranged with a parallel disposition of the loops within each β-sheet. This configuration has been proposed for the Q/N region of TDP-43, as fully described in Chapter 4.

These two structures were not obtained by atomic-resolution techniques, but built through computational modeling on the basis of experimental. The structure of the Aβ1-40 peptide carrying the Iowa mutation (D23N), determined to atomic resolution using ssNMR methods [85], is shown in **Fig. 13** to illustrate the information gathered by using full-length sequences. Among the eight possible cross-β spines, this structure has each strand in an antiparallel configuration with respect to the others. The loops are arranged in the β-arc topology, *i.e.*, each β-strand contributes to a different β-sheet, with an antiparallel configuration, as loops in consecutive monomers are not stacked.

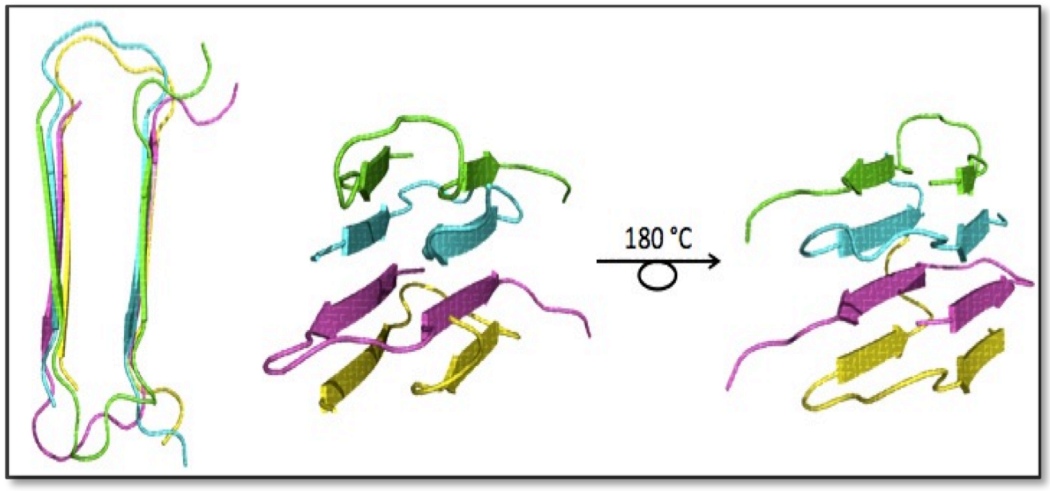


Fig. 13. Structure of the 40-residue Aβ peptide bearing the Iowa mutation (D23N). ssNMR structure reveals an antiparallel β-arc topology (β-loop-β, each strand contributes to one sheet). Each individual β-arc monomer is shown in a different color (yellow, pink, blue, and green). The topology is antiparallel since loops are oriented in different directions among consecutive monomers.

More than one cross-β spine can be found in mature fibrils **–** So far, in describing the cross-β spine features, we have only considered a pair of β-sheets. However, two or three of these "proto-fibrils" can entwine to form high-order arrangements in mature fibrils. Although this can be inferred from scanning-transmission electron microscopy (STEM), as well as from fibril's estimated widths observed under transmission electron microscopy (TEM), solid-state NMR has afforded direct observation of these arrangements at high resolution in Aβ fibrils, [86] as shown in **Fig. 14**. This highlights solid-state NMR's promise for studying amyloid fibrils.

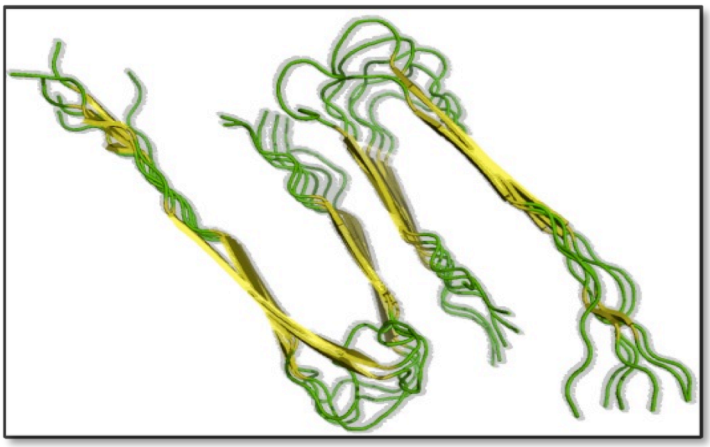


Fig. 14. ssNMR 3D-structure of Aβ fibrils showing association of two cross-β spines. Spectrally unambiguous contacts define an antiparallel arrangement that can be regarded a dimer of two cross-β spine units.

Nuclear Magnetic Resonance expands the amyloid foldome **–** New developments in solid-state NMR rotors and pulse sequences have allowed the determination of an increasing number of amyloid structures, highlighting some differences with respect to the canonical cross-β spine defined on the basis of available X-ray models. This powerful tool has an important advantage: it affords the study of full-length proteins with exclusive detection of the more rigid, static amyloid cores. This has revealed new structural features, such as novel arrangements of the monomers in solenoid [29] cinnamon-roll [87] or U-shaped [88] geometries (**Fig. 15A, 15B, 15C**).
ssNMR has also broadened our knowledge of amyloid polymorphism, since fibrils can be prepared in more physiological-like conditions without artificial crystal packing. As an example, a three-fold symmetry polymorph was observed for the Aβ fibrils (**Fig. 15D**).

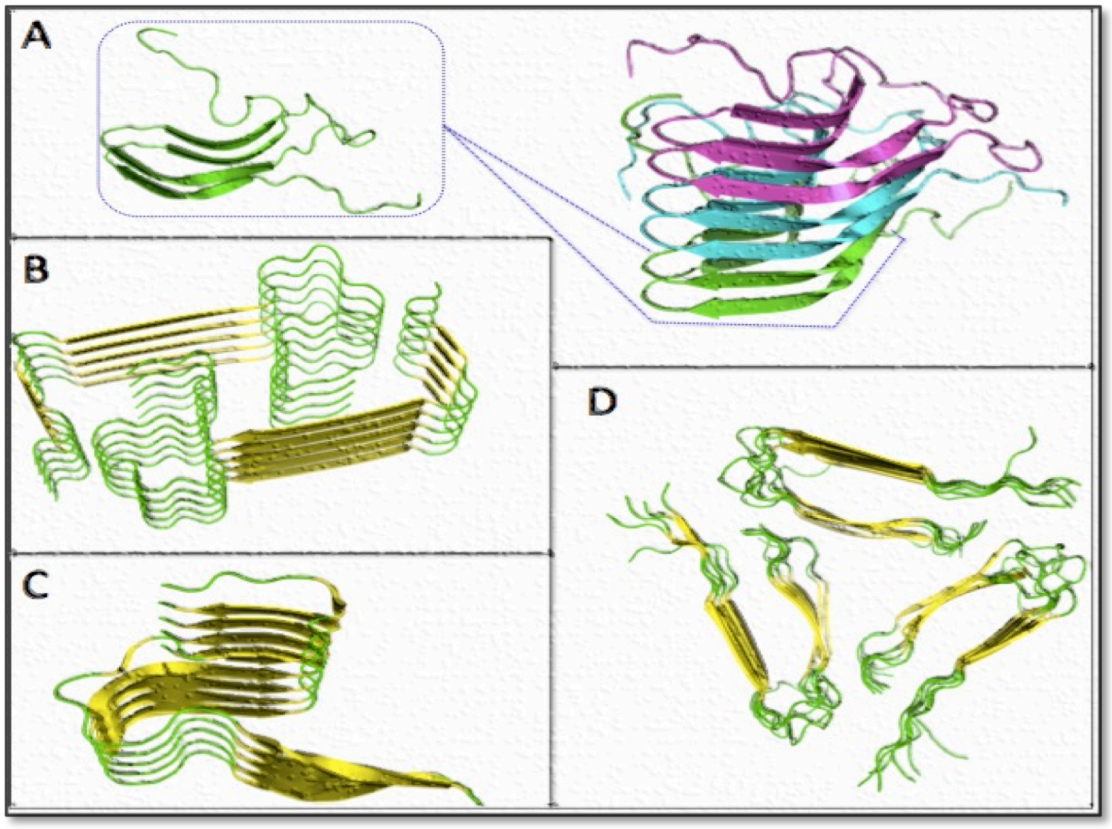


Fig. 15. **Solid-state NMR structures of amyloid fibrils.** (**a**) The HET-S functional prion adopts a β-solenoid architecture. The basic, repetitive unit is shown in green alone (left) and with two additional solenoids (blue and pink) forming a fibril. (**b**) Aβ1-40 bearing the Osaka mutation adopts a cinnamon-roll-like topology that differs from canonical cross-β spine. (**c**) First reported structure of the Aβ1-42 fibrils build on the basis of experimental distance constraints obtained from selectively labeled samples by solid-state NMR. (**b**) Aβ1-40 fibrils showing a three-fold symmetry

Amyloid polymorphism: does it need to be avoided? **–** In preparing samples for ssNMR, great care is generally exercised to samples prepared under the same exact conditions to avoid polymorphic structures that might hamper the spectral analysis required to achieve a high-resolution structure. This obsession to avoid polymorphism may wane, since recent results suggest that polymorphic structures may correlate with variations in disease prognoses. In fact, it has been shown that different Aβ assemblies exert different effects at the biological level, causing membrane disruption [89], generating reactive oxygen species [90], or by induction of inflammation [91]. Even in the previously described functional amyloid Pmel17, solid-state NMR has revealed the existence of different polymorphs under physiological conditions [92]. All these findings indicate that it might be necessary to solve all the possible structures that might arise from the same amyloidogenic proteins. It is fascinating that in 2013, Tycko's laboratory demonstrated that seeding Aβ1-40 fibrils from two AD patients' brain tissue resulted in a single predominant polymorph for each sample, but with structural variations between the two patients, which may correlate with variations in disease progression [93].

Despite the increasing knowledge on amyloids that is being gathered on the basis of new high-resolution structures and biophysical studies, their mechanism(s) of formation remain elusive. The main unsolved issue is how a monomer undergoes the structural transition into the β-sheet oligomeric state. Whereas there are experimental techniques, such as dynamic light scattering or fluorescence spectroscopy using small amyloid-binding molecules,,that allow the study of the fibrillization process, the structural transition itself can be tracked by circular dichroism spectroscopy, as it is illustrated in Chapter 3. However, it is not yet possible to follow amyloid formation at atomic resolutions in real time.

1.3 Transactive Response DNA-binding Protein 43 kDa (TDP-43)

TAR DNA-binding protein 43 kDa (TDP-43) is a 414-residue RNA-binding protein that belongs to the heterogeneous nuclear ribonucleoprotein (hnRNP) family, and is highly conserved within the animal kingdom. TDP-43 binds UG-repeated motifs of single-stranded RNAs [94, 95] and either silences or enhances splicing. Two representative examples of this antagonistic behavior are the exon skipping in the case of the cystic fibrosis transmembrane regulator (CFTR) exon9 [96], and the increased splicing in a breast cancer 1 (BRCA1)-mutated substrate [97], which are illustrated in the next **Fig. 16**.

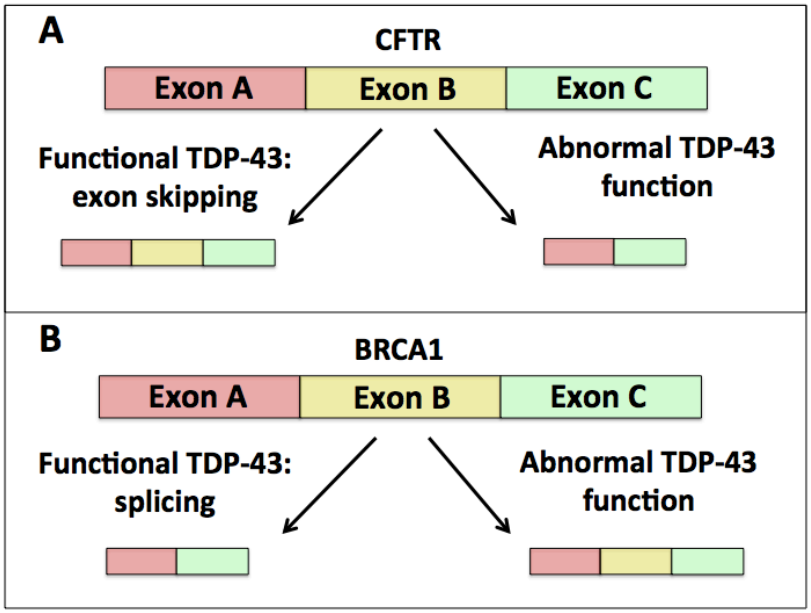


Fig. 16. Schematic representation of TDP-43 functions in regulation of splicing processes. **(A)** A hypothetic mRNA segment with three exons is shown in red (exon A), yellow (exon B) and green (exon C). Functional TDP-43 inhibits alternative splicing and promotes exon skipping, such that the normal function is altered, splicing occurs and exon B is removed. This abnormality could be due to formation of aggregates or by depletion of nuclear levels of the protein. **(B)** Although less common, it could be that TDP-43 acts enhancing splicing and therefore, skipping of exon B would be an abnormal function in this case.

To carry out these functions, TDP-43 has two RNA recognition motifs (RRMs) responsible for binding nucleic acids [98] and sequences for nuclear localization and exporting, which allow this protein to shuttle between the nucleus and the cytoplasm to mediate mRNA transportation [99]. Mutation of these sequences leads to the formation of aggregates and abnormal protein redistribution within the nucleus or the cytoplasm [100]. The protein also contains a well-folded N-terminal domain [101], described in Chapter 5, and a Gly-rich C-terminal domain [102], which has been shown to contain a prion-like domain rich in Asn and Gln residues [103, 104], as reported in Chapters 3 and 4. TDP-43's domain organization and sequence are shown in **Fig. 17**.

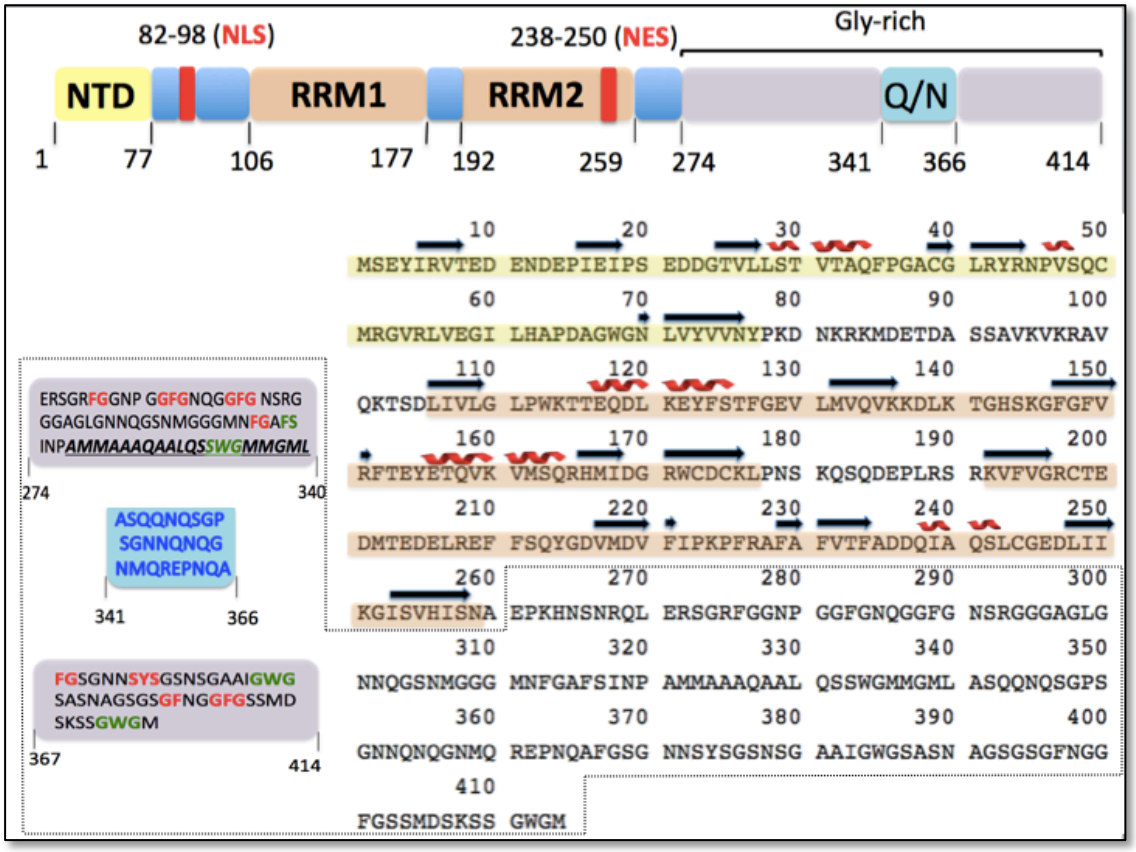


Fig. 17. Structural organization of TDP-43 domains. TDP-43 contains two RNA binding domains (106-177 & 192-259, orange) for nucleic acid binding. Since it shuttles in and out of the nucleus, both nuclear localization and export sequences (NLS & NES) are present in the protein (residues 82-98 and 238-250, respectively). These regions are shown in red in the figure. The N-terminal domain, whose 3D structure is reported in Chapter 5, is composed by the first 77 residues (yellow). The large C-terminal region contains a Gly-rich region (purple) and prion-like segment (residues 341-366, colored in light blue). It also contains a hydrophobic patch composed of residues 321-340 (bolds, italics & underscored) and LC sequences, which are shown in red and green, according to the color code used in previous Fig. 8. The secondary structure elements that have been solved at high-resolution are displayed in black (β-strands) or red (α-helices).

TDP-43 is attracting increased attention since it was found to be the major component in ubiquitin-positive, tau-, and alpha-synuclein-negative proteinaceous inclusions in patients with ALS and FTLD-TDP diseases [30]. Although other proteins such as SOD, FUS/TLS, and dipeptide repeat proteins have been implicated in some cases of ALS [105–111], TDP-43 inclusions are seen in over 95% of ALS cases, which strongly suggests that it is probably the major culprit in this disease.

A minority of ALS cases linked to TDP-43 has a genetic component and almost all of the disease-related mutations are located in the C-terminal region of TDP-43 [112, 113]. It is notable that the pathological aggregates contain 20-25 kDa C-terminal fragments (CTFs) with an increased potential for aggregation and cytotoxicity [114].

This, and the fact that the C-terminal region mediates interactions with other hnRNPs [115] suggests that this domain of TDP-43 plays a key role in the protein functionality and consequently, in ALS and FTLD pathogenesis. Moreover, the C-terminal prion-like segment is able to disrupt endogenous TDP-43—hnRNP A2 complexes [115], and a construct bearing twelve tandem repeats of this segment attached to the C-terminus of TDP-43 is able to efficiently recruit the endogenous WT protein into aggregates [116], which then undergoes the post-translational modifications (phosphorylation and ubiquitination) observed in the disease state [117]. This depletion of TDP-43 levels in the nucleus causes a loss-of-function at the mRNA splicing level, and the presence of the N-terminal domain (NTD) is essential to promote the interaction of this construct with the endogenous protein [118]. In addition to this role in pathogenesis, the NTD has been found to be essential for homodimer formation and normal splicing functions [119]. For these reasons, the Q/N and NTD domains are widely studied in Chapters 3, 4, and 5.

It is interesting that TDP-43 proteinopathies have been documented in other neuronal disorders, such as those caused by Aβ or α-synuclein inclusions [120, 121], but probably as a secondary event since the amount and distribution of TDP-43 inclusions is minor. In fact, TDP-43 oligomers have been shown to inhibit Aβ40 fibrillization towards spherical oligomers formation *in vitro* [31].

1.4 Route map of this thesis

This section provides a comprehensive link between the Chapters presented in this thesis. Following the previous introduction aimed at highlighting the widespread existence of amyloids in diverse organisms, their fold and pernicious or functional nature, and after introducing TDP-43, a Chapter-by-Chapter description is given now to emphasize the scope and relationships between the next Chapters.

Considering that amyloidogenesis is regarded a nucleation process, Chapter 2 addresses the formation of the initial seed that triggers aggregation using a model system: the heptapeptide GNNQQNY from the prion yeast protein Sup35, whose structural features have already been introduced. Extensive computational calculations were performed, and the empirical parameters used in the simulations were successfully validated as will be described. The reason to choose theoretical approaches and computational techniques is based on the fact that current experimental methods do not yet allow for a precise characterization of amyloid formation. Amyloid-forming peptides that are present in solution in a disordered conformation can be characterized by NMR spectroscopy. However, when aggregation is triggered the NMR signal intensity decreases as larger oligomers tumble too slowly to permit their observation (**Fig 18**).

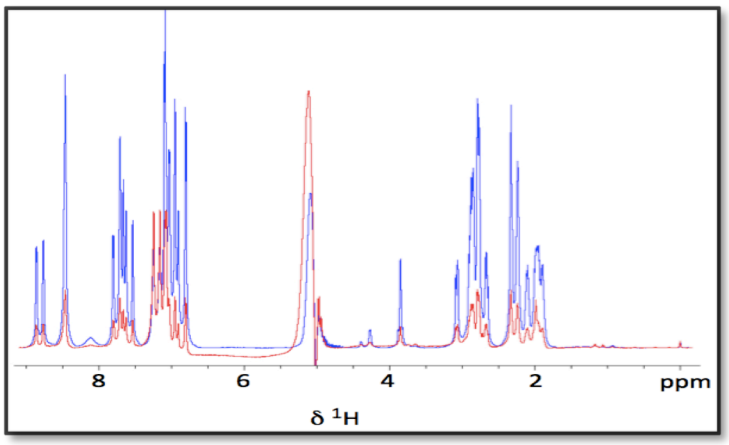


Fig. 18. ^{1}H-NMR to monitor peptide aggregation. Solution NMR can be used to detect the conformations prior to β-sheet conversion upon amyloid formation. In the graph, the spectra of freshly dissolved Sup35 heptapeptide (blue) and after five days (red) are shown. Large oligomer and aggregate formation reduces the amount of monomeric peptide detected by solution state NMR.

On the other hand, solid-state NMR is very suitable to characterize the aggregates, but it cannot provide information about the monomer that once was present in solution. The same applies for X-ray diffraction, as it can only provide a static picture of fibrils or crystals. In contrast, CD spectroscopy can follow secondary structure transitions in real time, but cannot provide structural details at atomic resolution. Therefore, computational techniques have been widely applied to gain insight into this conversion at an atomistic scale. Here, a combination of experimental techniques and computation is done to go beyond the theoretical picture and validate the computational results.

The same methodology used in Chapter 2 for the GNNQQNY fibril was then applied in Chapter 3 to develop a model for the monomeric, aggregation-prone conformation of an amyloid-forming peptide from the prion-like Gln/Asn-rich segment of the C-terminal region of TDP-43. This model was built on the basis of data obtained from NMR, CD, and fluorescence spectroscopies. Chapter 4 builds on the monomeric structure developed in Chapter 3 to generate models for amyloid-like oligomers, which are supported by a broad battery of biophysical techniques. Finally, Chapter 5 is dedicated to the elucidation at high-resolution of the 3D structure, conformational stability and backbone dynamics of the N-terminal domain.

The contributions of the Q/N region and the N-terminal domain in the ability of TDP-43 to regulate splicing, homodimerize, and aggregate have been previously described. Therefore, these results contribute with special interest to our knowledge of TDP-43 functions and pathology for two main reasons: i) the amorphous vs. amyloid-like nature of key aggregates is clarified, and ii) the highly stable, distinct fold showed by the NTD, which contrasts with the recent proposal of its direct implication in the pathological setting through a ubiquitin-like, marginally stable structure implicated in nucleic acid binding [122]. All these findings are combined into a model for full-length TDP-43 that intends to be fully compatible with the experimental data currently found in the literature, and that can be updated as new data becomes available.

The work presented in Chapters 2, 3, 4, and 5 has been published. In Chapter 6, I describe two research lines that are currently in progress:

1. Unifying the structural models of the N-terminal, RNA binding domains, and C-terminal region into a structural model for full-length TDP-43.
2. Interactions that stabilize amyloids rich in polar residues (PolyQ, Sup35,TDP-43…) versus those rich in hydrophobic residues, like Aβ.

Finally, Chapter 7 contains the conclusions derived from the work performed during these three years of research and, they are not included in this *Route Map* section.

1.4.1 From the random-coil ensemble towards the initial seed for amyloid growth

Amyloids are believed to form through a nucleation/growth mechanism, in which the rate-limiting step is the slow assembly of a minimal oligomer, which then grows quickly. It was previously described that the assembly of two β-sheets to yield a dry interface and mature amyloid fibrils occurs through interdigitation of the residue's side chains, providing a set of van der Waals interaction at this contact area devoid of water. This is not surprising for residues with hydrophobic sidechains where these non-polar contacts are strongly stabilizing. However, the case of GNNQQNY is of particular interest since highly polar side chains are the driving force for two β-sheets to fold into a cross-β spine.

Given the premise that Gln- and Asn-rich sequences drive amyloid formation in prions, we decided to use the GNNQQNY as a model system to study in detail the initial nucleus displaying amyloid structural features. This system is an excellent candidate for these investigations, as its prion-like behavior resembles the properties of the full-length protein [123], while allowing for precise *Ab-initio* Quantum Mechanical calculations and long simulation times through classical Molecular Dynamics due to its short length. In this work, which is presented in Chapter 2, we proposed a model for nucleation of amyloid fibers, characterizing the minimal aggregation core required for fibril elongation (**Fig. 19**).

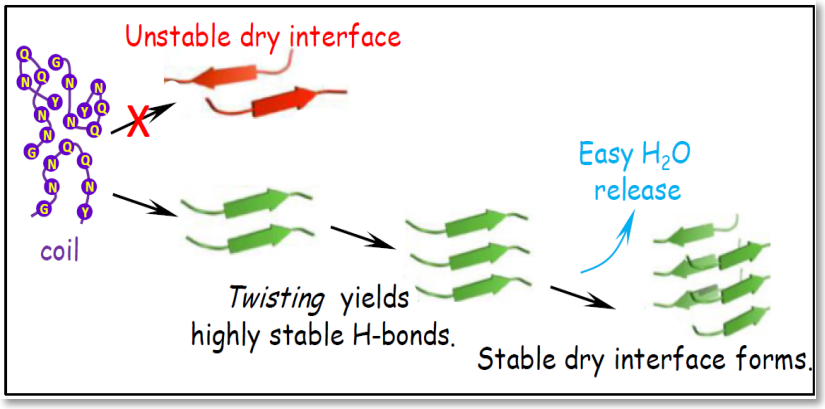


Fig. 19. **Graph summarizing results presented in Chapter 2.** Sup35 evolves from a chiefly disordered ensemble into a parallel, three-stranded β-sheet that is ready to pack and form a cross-β spine without mediation of water nanowires. Dry interface formation requires at least that one β-sheet of this length, and twisting changes the hydrogen bonding pattern of this structural motif and provides stability.

Previous work on the Sup35 heptapeptide led to the proposal that the formation of two β-sheets and their mutual approach would be relatively rapid, trapping a layer or "nanowire" of water molecules at the dry interface. The rate-limiting step for amyloid formation would then correspond to the slow release of these trapped water molecules (**Fig. 20**). It was proposed that water molecules could surround the polar residues in GNNQQNY to mediate assembly of the β-sheets through long-lived water nanowires [124]. This is not likely to occur in hydrophobic segments and therefore, the hypothesis of desolvation as the driving force for cross-β spine formation in polar sequences gain reliability [125]. In fact, a similar study on amyloid fibrils from the human amylin showed very similar results [126]. Nevertheless, both these two studies used long β-sheets each containing eight β-strands, and both lacked an experimental validation of the force field. Here, I show that β-sheet as short as composed of three strands are readily able to pack against each other without stable water nanowires formation, along with a validation of the set of parameters used in the simulations.

Prior to our work, the ability of smaller β-sheets to pack without water mediation were not previously tested. Therefore, our results are important in light of a proposed two-step mechanism for amyloid aggregation in which water solvation, prior to dry interface formation, might contribute to amyloid polymorphism [125]**.** Moreover, our model also explains the advantages of twisted amyloid fibrils with a rigorous Quantum Mechanics formalism.

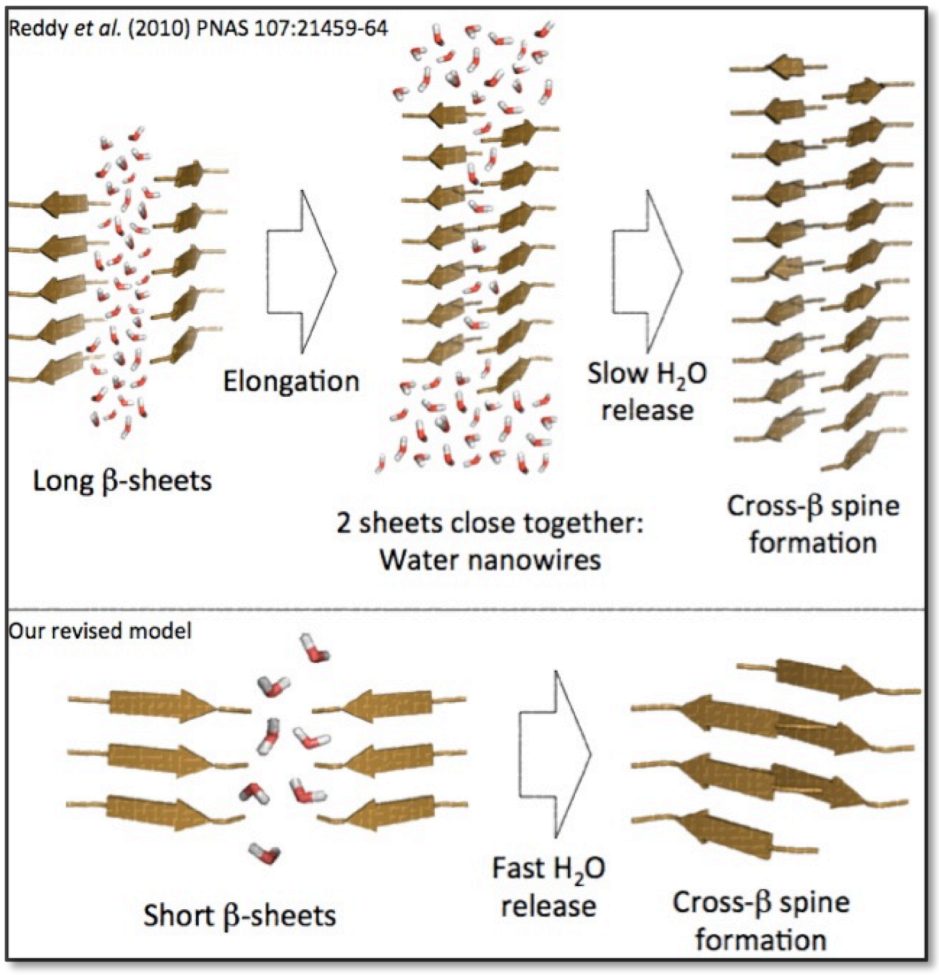


Fig. 20. **Water nanowires were proposed to mediate fibril assembly.** It was proposed that water molecules separating the mating β-sheets hamper dry interface formation. As the two sheets come close together to pack together, formation of water nanowires persisting for *ca*. 20 ns would hinder the assembly. Although this possibility seemed reasonable, the β-sheets considered in that work were too long, and we show in Chapter 2 that shorter sheets can readily packed much more quickly with facile desolvation.

Finally, we performed solution NMR experiments to validate the theoretical parameters used in this work by means of a peptide in which the GNNQQNY sequence is intentionally scrambled to impede oligomerization. MD simulations successfully predicted this peptide's inability to aggregate, which was proved by fluorescence and NMR spectroscopies. Interestingly, a 1:1 mixture of the WT and the scrambled peptide results in inhibition of amyloid formation (**Fig. 21**).

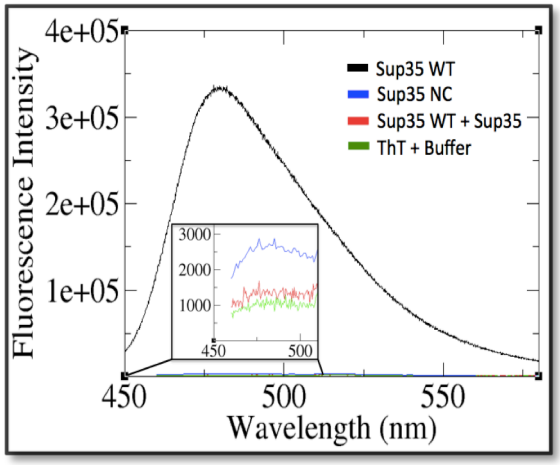


Fig. 21. **ThT assays on GNNQQNY and the scrambled peptide.** ThT fluorescence enhancement indicated amyloid formation in the Sup35 peptide GNNQQNY (black curve), but not in the case of the scrambled sequence used as a negative control (NC, in blue), nor in a 1:1 mixture of both (red), suggesting inhibition of aggregation. A blank (ThT + buffer) is shown in green.

1.4.2 Application to a specific system: elucidation of the minimal segment of TDP-43 that is competent for aggregation

The methodologies applied before in Chapter 2 with GNNQQNY showed that Amber99SB-ildn provides a satisfactory set of parameters that describe the behavior of amyloid-forming peptides containing polar residues fairly well. Therefore, we utilized this force field to characterize an aggregation-prone segment within the C-terminal domain of TDP-43.

Although some studies have dealt with amyloid formation in the hydrophobic segment 321-340 [127, 128], the prion-like region 341-366 was shown to be essential to induce aggregation in cells, as well as competition for the interaction with other hnRNPs. Bearing this in mind, we decided to characterize the structural features of peptides that correspond to the Gln/Asn-rich segments within the full-length TDP-43. We used CD, NMR and MD to characterize a structural conversion from a random coil ensemble into an aggregating β-hairpin oligomer, as shown in **Fig. 22**.

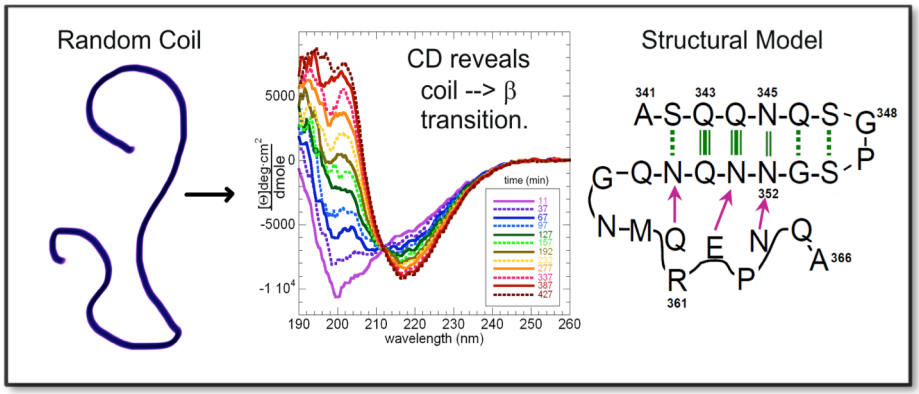


Fig. 22. Graphic summarizing results presented in Chapter 3. The Q/N-rich, prion-like segment of TDP-43's C-terminal domain is initially in a random coil conformation. CD spectroscopy allows to track the conversion into a β-rich oligomer that aggregates. MD simulations were used to model this segment's possible conformations based on spectroscopy data from the wt segment and some mutants. The preferred conformation that we found is a β-hairpin (341-357) possibly followed by a structured, yet not canonical segment (358-366).

1.4.3 Thorough characterization of TDP-43(341-357) aggregates

Our work with the peptide corresponding to the 341-366 segment suggested that a smaller peptide consisting of just the residues forming the β-hairpin (341-357) would be sufficient to form amyloids. The results presented in Chapter 4 illustrate how this segment could oligomerize, as well as the nature of the aggregates and the basis of their conformational stability. Some pathological mutations were also studied, and we found that peptides carrying these substitutions do not aggregate.

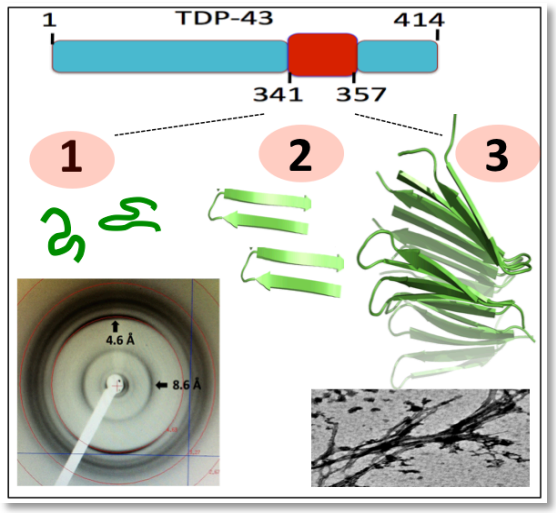


Fig. 23. Graph summarizing results presented in Chapter 4. The segment characterized in Chapter 3 as competent for aggregation has amyloid like properties. TDP-43 residues 341-357 yield a cross-diffraction pattern typical of amyloid fibrils. This figure illustrates schematically how monomers without preferred conformations form β-hairpins and then oligomerize into amyloid fibrils. An electron microscopy image of these fibrils is shown.

The amorphous versus amyloid-like nature of the C-terminal inclusions formed by TDP-43 has been debated in the literature [128, 129]. As advanced in the previous section, we found evidence for the formation of an oligomer rich in β-structure by the Gln/Asn-rich region. This finding prompted us to extensively characterize the possible amyloid nature of those aggregates, and as suggested by the results of previous Thioflavin-T assays. We also recorded a cross diffraction pattern typical of amyloid fibrils, but considering that in rare cases coiled coils have yielded similar reflections [81], we carried out an extensive set of experiments including solid-state NMR, recognition by conformational antibodies, electron microscopy and circular dichroism. The results of these tests, reported in Chapter 4, show that the peptide corresponding to TDP-43(341-357) forms amyloid-like fibrils. Based on these findings, we utilized a variety of computational strategies to build structural models for TDP-43(341-357) oligomers in which β-hairpins assemble into a novel parallel β-turn configuration that expands the amyloid foldome (**Fig. 23**). We will show in Chapter 6 that this model is consistent with the conformation of the full-length protein and with results from *in-cell* experiments.

1.4.4 High-Resolution structure of the NTD of TDP-43 implicated in aggregation

In a cellular context and considering the full-length protein, the N-terminal domain (NTD) of TDP-43 (residues 1-77) is essential to effectively recruit TDP-43 monomers into aggregates, presumably formed by the Q/N-rich segment described in previous sections. Given the tight link between this NTD with human biology and pathology, a high-resolution structure is essential to advance our knowledge of TDP-43.

While our investigations were underway, Qin *et al*. reported a low-resolution model of this domain on the basis of a small number (78) of NOE-derived distance restraints [122]. They showed that this domain, which would adopt a ubiquitin-like fold, was in a *folded* ←→ *unfolded* equilibrium, with the *folded* state stabilized by DNA binding. While this work advanced our understanding of TDP-43 functionality and pathological implications, its low stability and purported ability to bind DNA, which contradicts previous results [130], were puzzling.

Therefore, we continued our work. We show in Chapter 5 that this domain adopts a well defined, rare fold, on the basis of > 1000 NOE-derived distance restraints (**Fig. 24**), along with H/D exchange data that provided 26 H-Bonds restraints and revealed a conformational stability of 4 kcal/mol. This structure is very similar to the DIX domain of Axin1, which mediates protein-protein interactions [131]. The analysis of backbone dynamics on the basis of the ^{15}N-^{1}H heteronuclear NOE emplying the MODELFREE approach is in agreement with the structural features observed: six β-strands and an α-helix, where only one edge strand along with the loops are shown to be flexible. In Qin's work, the analysis of the ^{15}N-^{1}H heteronuclear NOE does not seem to

be in full agreement with their low-resolution model. Moreover, they used a longer construct with a net charge of ~ +14 for DNA binding assays, whereas in our stable NTD construct, which spans residues 1-77, there are two clusters of negatively charged residues. On this basis, we advance that nucleic acid binding might not be a function of the NTD.

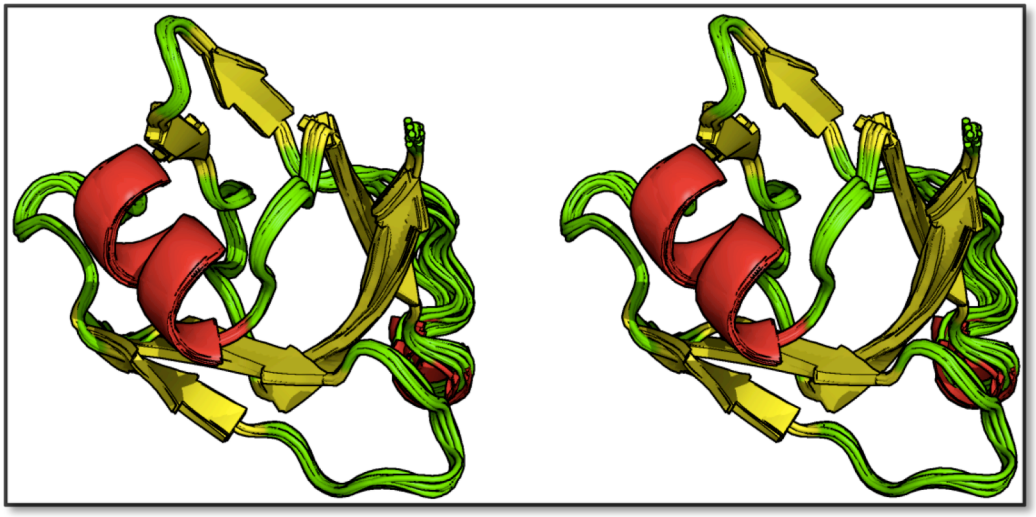


Fig. 24. Cross-eye stereoscopic view of the NMR structure of the NTD domain. The 20 lowest-energy conformers for the NMR structure of the N-terminal domain of TDP-43 are shown. The secondary structural elements are colored in yellow (β-strands) and red (α-helix), whereas other residues are shown in green. This structure is described in detail in Chapter 5, along with backbone dynamics.

1.4.5 Ongoing work: integrative model of full-length TDP-43 and amyloid energetics

Production of recombinant full-length TDP-43 is a challenging task due to its notorious sub-micromolar insolubility, and this has hampered obtaining the structure of this protein. Fortunately, elucidation of the structures of both the two RRM domains in solution has been recently achieved [132].
As stated in previous sections, we used NMR to obtain for the first time the structure of the N-terminal domain (residues 1-77) at high-resolution, and we have proposed a molecular model for the Q/N-rich region. These findings prompted us to combine all these results into a model of the full-length protein, which is intended to serve as a framework to address the essential questions regarding TDP-43's functions and implications in pathology.

Another ongoing project described in that Chapter is the study of basic aspects of amyloid formation. The motivation is that, as it has been described in this Introduction, different sequences give rise to the same fold with both detrimental effects and beneficial functions in humans, thwarting rational drug design aimed at disrupting amyloid formation. We have identified a ubiquitous property at the quantum mechanical level that does not depend on the length of the sequences or the residue composition, and provides the amyloid fold with its remarkable stability. We call it *residual Hydrogen Bonding Cooperativity* (re-HBC) and interestingly, Q/N-rich sequences contain a second term that is modulated by polar solvents capable of forming H-bonds with the fibrils. We denote this second contribution to amyloid stability as *hyper Hyrogen Bonding Cooperativity* (*hy-HBC*). This energetic treatment distinguishes the energetics of polar and hydrophobic amyloids and is supported by experimental measurements obtained with broadband dielectric spectroscopy. These findings correlate well with our observations by fluorescence and NMR spectroscopies and may help guide the future design of selective amyloid inhibitors.

1.4 References

(1) Pepys, M. B. Amyloidosis. *Annu. Rev. Med.* **2006**, *57*, 223-241.

(2) Alzheimer, A. *Allgemein Zeitschrift für Psyciatrie und Psychisch Gerichtliche Medizin* **1906**, *64*, 146-148.

(3) Glenner, G. G.; Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem. Biophys. Res. Commun.* **1984**, *120*, 885-890.

(4) Lambert, M. P.; Barlow, A. K.; Chromy, B. A.; Edwards, C.; Freed, R.; Liosatos, M.; Morgan, T. E.; Rozovsky, I.; Trommer, B.; Viola, K. L.; Wals, P.; Zhang C.; Finch, C. E.; Krafft, G. A.; Klein, W. L. Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6448-6453.

(5) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid and human disease. *Annu. Rev. Biochem.* **2006,** *75*, 333-336.

(6) Eisenberg, D.; Jucker, M. The amyloid state of proteins in human diseases. *Cell* **2012,** *148*, 1118-1203.

(7) Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W. Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.* **2007**, *32*, 217-224.

(8) Si, K.; Choi, Y. B.; White-Grindley, E.; Majumdar, A.; Kandel, E. R. Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell* **2010,** *140*, 421-435.

(9) Goldgaber, D.; Lerman, M. I.; McBride, O. W.; Saffiotti, U.; Gajdusek, D. C. Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease. *Science* **1987**, *235*, 877-880.

(10) Masters, C. L.; Multhaup, G; Simms, G.; Pottgiesser, J.; Martins, R. N.; Beyreuther, K. Neuronal origin of a cerebral amyloid: neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels. *EMBO J.* **1985**, *4*, 2757–2763.

(11) Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **2001,** *81*, 741-766.

(12) Selkoe, D. J. Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nat. Cell Biol.* **2004,** *6*, 1054-1061.

(13) Ikeda, K.; Tomonaga, M. Alzheimer's disease amyloid precursor protein is present in the myelin sheath of central nervous system in rat. *Brain Res*. **1990**, *527*, 140-144.

(14) Le, Y.; Gong, W.; Tiffany, H. L.; Tumanov, A.; Nedospasov, S.; Shen, W.; Dunlop, N. M.; Gao, J. L.; Murphy, P. M.; Oppenheim, J. J.; Wang, J. M. Amyloid (beta)42 activates a G-protein-coupled chemoattractant receptor, FPR-like-1. *J. Neurosci*. **2001,** *21*, RC123.

(15) Soscia, J. S.; Kirby, J. E.; Washicosky, K. J.; Tucker, S. M.; Ingelsson, M.; Hyman, B.; Burton, M. A.; Glodstein, L. E.; Duong, S.; Tanzi, R. E.; Moir, R. D. The Alzheimer's disease amyloid β-protein is an antimicrobial peptide. *PLoS One*. **2010,** *5*, e9505.

(16) Domert, J.; Rao, S. B.; Agholme, L.; Brorsson, A. C.; Marcusson, J.; Hallbeck, M.; Nath, S. Spreading of amyloid-β peptides via neuritic cell-to-cell transfer is dependent on insufficient celular clearance. *Neurobiol. Dis*. **2014,** *65*, 82-92.

(17) Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Dobeli, H.; Schubert, D.; Riek, R. 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proc. Natl. Acad. Sci. U. S. A*. **2005,** *102*, 17342-17347.

(18) Chandra, S.; Chen, X.; Rizo, J.; Jahn, R.; Südhof, T. C. A broken alpha-helix in folded alpha-Synuclein. *J. Biol. Chem*. **2003,** *278*, 15313-15318.

(19) Conway, K. A.; Lee, S. J.; Rochet, J. C.; Ding, T. T.; Williamson, R. E.; Lansbury, P. T. Jr. Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. U. S. A*. **2000,** *97*,571-576.

(20) Lansbury, P. T. Jr.; Brice, A. Genetics of Parkinson's disease and biochemical studies of implicated gene products. *Curr. Opin. Genet. Dev*. **2002,** *12*, 299–306.

(21) Kirkitadze MD, Bitan G, Teplow DB (2002) Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *J Neurosci Res* **2002,** *69,* 567–577.

(22) Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **2002,** *416,* 507–511.

(23) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **2003,** *300,* 486-489.

(24) Hervás, R.; Oroz, J.; Galera-Prat, A.; Goñi, O.; Valbuena, A.; Vera, A. M.; Gómez- Sicilia, A.; Losada-Urzáiz, F.; Uversky, V. N.; Menéndez, M.; Laurents, D. V.; Bruix, M; Carrión-Vázquez, M. Common features at the start of the neurodegeneration cascade. *PLoS Biol.* **2012,** *10*, e1001335.

(25) McGowan, D. P.; van Roon-Mom, W.; Holloway, H.; Bates, G. P.; Mangiarini, L.; Cooper, G. J.; Faull, R. L.; Snell, R. G. Amyloid-like inclusions in Huntington's disease. *Neuroscience* **2000**, *100*, 677-680.

(26) Petruska, J.; Hartenstine, M. J.; Goodman, M. F. Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *Journal of Biological Chemistry* **1998,** *273,* 5204–5210.

(27) Velier, J.; Kim, M.; Schwarz, C.; Kim, T. W.; Sapp, E.; Chase, K.; Aronin, N.; DiFiglia, M. Wild-type and mutant huntingtins function in vesicle trafficking in the secretory and endocytic pathways. *Exp. Neurol.* **1998,** *52,* 34-40.

(28) Cummings, C. J.; Zoghbi, H. Y. Fourteen and counting: unraveling trinucleotide repeat diseases. Hum. Mol. Genet. **2000,** *9,* 909-916.

(29) Albrecht, A.; Mundlos, S. The other trinucleotide repeat: polyalanine expansion disorders. *Curr. Opin. Genet. Dev.* **2005,** *15,* 285-293.

(30) Neumann, M.; Sampathu, D. M.; Kwong, L. K.; Truax, A. C.; Micsenyi, M. C.; Chou, T. T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C. M.; McCluskey, L. F.; Miller, B. L.; Masliah, E.; Mackenzie, I. R.; Feldman, H.; Feiden, W.; Kretzschmar, H. A.; Trojanowski, J. Q.; Lee, V. M. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006,** *314,* 130-133.

(31) Fang, Y. S.; Tsai, K. J.; Chang, Y. J.; Kao, P.; Woods, R.; Kuo, P. H.; Wu, C. C.; Liao, J. Y.; Chou, S. C.; Lin, V.; Jin, L. W.; Yuan, H. S.; Cheng, I. H.; Tu, P. H.; Chen, Y. R. Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* **2014**, *5,* 4824.

(32) Chapman, M. R.; Robinson, L. S.; Pinkner, J. S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S. K. Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **2002,** *295*, 851-855.

(33) Shewmaker, F.; McGlinchey, R. P.; Wickner, R. B. Structural insights into functional and pathological amyloid. *J. Biol. Chem.* **2001,** *286,* 16533-16540.

(34) Blanco, L. P.; Evans, M. L.; Smith, D. R.; Badtke, M. P.; Chapman, M. R. Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol.* **2012,** *20,* 66-73.

(35) True H. L.; Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **2000,** *407,* 477-483.

(36) Shorter, J.; Lindquist, S. Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers. *Science*, **2004,** *304,* 1793-1797.

(37) Jarosz, D. F.; Jones, S. K.; Chang, A.; Lancaster A. K.; Lindquist, S. Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature* **2012,** *482,* 363-368.

(38) Coustou, V.; Deleu, C.; Saupe, S.; Bequeret, J. The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94,* 9773-9778.

(39) Ritter, C; Maddelein, M. L.; Siemer, A. B.; Lührs, T.; Meier, B. H.; Saupe, S. J.; Riek, R. Correlation of structural elements and infectivity of the HET-s prion. *Nature* **2005,** *435,* 844-848.

(40) Wasmer, C.; Lange, A.; van Melckebeke, H.; Siemer A., B.; Riek, R.; Meier B. H. Amyloid fibrils of the HET-s (218-289) prion form a β solenoid with a triangular hydrophobic core. *Science* **2008,** *319,* 1523-1526.

(41) van Melckebeke, H.; Wasmer, C.; Lange, A.; Eiso, A.B.; Loquet, A.; Böckmann, A.; Meier, B. H. Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state nmr spectroscopy. *J. Am. Chem. Soc.* **2010,** *132,* 13765-13775.

(42) Balguerie, A.; Dos Reis, S.; Ritler, C.; Chaignepain, S.; Coulary-Salin, B.; Forge, V.M Bathany, K.,; Lascu, I., Schmitter, J. M Riek, R.; Saupe, S. J. Domain organization and structure–function relationship of the HET-s prion protein of *Podospora anserina*. *EMBO J.* **2003,** *22,* 2071-2081.

(43) Glass, N. L.; Jacobson, D. J.; Shiu, P. K. The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annu. Rev. Genet.* **2000,** *34,* 165-186.

(44) Saupe, S. J. Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiol. Mol. Biol. Rev.* **2000,** *64,* 489-502.

(45) Turcq B.; Deleu,C.; Denayrolles, M.; Bégueret, J. Two allelic genes responsible for vegetative incompatibility in the fungus *Podospora anserina* are not essential for cell viability. *Mol. Gen. Genet.* **1991,** *228,* 265–269.

(46) Berthelot, K.; Lecomte, S.; Estevez, Y.; Coulary-Salin, B.; Bentaleb, A.; Cullin, C.; Deffiewux, A.; Peruch, F. Rubber elongation factor (REF), a major allergen component in *Hevea brasiliensis* latex has amyloid properties. *PLoS One* **2012,** *7,* e48065.

(47) Watt, B.; van Niel, G.; Raposo, G.; Marks, M. S. PMEL: a pigment cell-specific model for functional amyloid formation. *Pigment Cell Melanoma Res*. **2013,** *26,* 300-315.

(48) Berson, J. F.; Harper, D. C.; Tenza, D.; Raposo, G.; Marks, M. S. Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. *Mol. Biol. Cell* **2001,** *12,* 3451–3464.

(49) Marks, M. S.; Seabra, M. C. The melanosome: membrane dynamics in black and white. *Nat. Rev. Mol. Cell Biol*. **2001,** *2,* 738–748.

(50) Berson, J. F.; Theos, A. C.; Harper, D. C.; Tenza, D.; Raposo, G.; Marks, M. S. Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *J. Cell Biol*. **2003,** *161,* 521–533.

(51) Festjens, N.; Vanden Berghe, T.; Cornelis, S.; Vandenabeele, P. RIP1, a kinase on the crossroads of a cell's decision to live or die. *Cell Death Differ*. **2007,** *14,* 400–410.

(52) Declercq, W.; Vanden Berghe, T.; Vandenabeele, P. RIP kinases at the crossroads of cell death and survival. *Cell* **2009,** *138,* 229–232.

(53) Meylan, E.; Tschopp, J. The RIP kinases: crucial integrators of cellular stress. *Trends Biochem. Sci*. **2005,** *30,* 151-159.

(54) Li, J.; McQuade, T.; Siemer, A.B.; Napetschnig, J.; Moriwaki, K.; Hsiao, Y.S.; Damko, E.; Moquin, D.; Walz, T.; McDermott, A. E.; Chan, F. K.; Wu, H. The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **2012,** *150,* 339-350.

(55) Si, K.; Giustetto, M.; Etkin, A.; Hsu, R.; Janisiewicz, A. M.; Miniaci, M. C.; Kim, J. H.; Zhu, H.; Kandel, E. R. A neuronal isoform of CPEB regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in *Aplysia*. *Cell* **2003,** *115,* 893-904.

(56) Si K.; Choi, Y.B.; White-Grindley, E.; Majumdar, A.; Kandel, E. R. Aplysia CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell* **2010,** *140,* 421-43.

(57) Si K.; Lindquist S.; Kandel E. R. A neuronal isoform of the *Aplysia* CPEB has prion-like properties. *Cell* **2003,** *115,* 879-891.

(58) Raveendra B. L.; Siemer A. B.; Puthanveettil S. V.; Hendrickson W. A.; Kandel E. R.; McDermott A. E. Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. *Nat. Struct. Mol. Biol*. **2013,** *20,* 495-501.

(59) Majumdar, A.; Cesario, W. C.; White-Grindley, E.; Jiang, H.; Ren, F.; Khan, M. R.; Li, L.; Choi, E. M.; Kannan, K.; Guo, F.; Unruh, J.; Slaughter, B.; Si, K. Critical role of amyloid-like oligomers of *Drosophila* Orb2 in the persistence of memory. *Cell* **2012,** *148,* 515-529.

(60) White-Grindley, E.; Li, L.; Mohammad Khan, R.; Ren, F.; Saraf, A.; Florens, L.; Si, K. Contribution of Orb2A Stability in Regulated Amyloid-Like Oligomerization of *Drosophila* Orb2. *PLoS Biol*. **2014,** *2*, e1001786.

(61) Fioriti, L.; Myers, C.; Huang, Y. Y.; Li, X.; Stephan, J. S.; Trifilieff, P.; Colnaghi, L.; Kosmidis, S.; Drisaldi, B.; Pavlopoulos, E.; Kandel, E. R. The persistence of hippocampal-based memory requires protein synthesis mediated by the prion-like protein CPEB3. *Neuron* **2015**, *86*, 1433-1448.

(62) Iko, Y.; Kodama, T. S.; Kasai, N.; Oyama, T.; Morita, E. H.; Muto, T.; Okumura, M.; Fujii, R.; Takumi, T.; Tate, S.; Morikawa, K. Domain architectures and characterization of an RNA-binding protein, TLS. *J. Biol. Chem*. **2004,** *279,* 44834-44840.

(63) Lee, B. J.; Cansizoglu, A. E.; Suel, K. E.; Louis, T. H.; Zhang, Z.; Chook, Y. M. Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **2006,** *126,* 543-558.

(64) Zhang, Z. C.; Chook, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 12017-12021.

(65) Dormann, D.; Rodde, R.; Edbauer, D.; Bentmann, E.; Fischer, I.; Hruscha, A.; Than, M. E.; Mackenzie, I. R.; Capell, A.; Schmid, B.; Neumann, M.; Haass, CALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J*. **2010,** *29,* 2841-2857.

(66) Ticozzi, N.; Silani, V.; LeClerc, A. L.; Keagle, P.; Gellera, C.; Ratti, A.; Taroni, F.; Kwiatkowski, T. J. Jr.; McKenna-Yasek, D. M.; Sapp, P. C.; Brown, R. H. Jr.; Landers, L. E. Analysis of *FUS* gene mutation in familial amyotrophic lateral sclerosis within an Italian cohort. *Neurology* **2009,** *73,* 1180-1185.

(67) Nomura, T.; Watanabe, S.; Kaneko, K.; Yamanaka, K.; Nukina, N.; Furukawa, Y. Intranuclear aggregation of mutant FUS/TLS as a molecular pathomechanism of amyotrophic lateral sclerosis. *J. Biol. Chem*. **2014,** *289,* 1192-1202.

(68) Kato, M.; Han, T. W.; Xie, S.; Shi, K.; Du, X.; Wu, L. C.; Mirzaei, H.; Goldsmith, E. J.; Longgood, J.; Pei, J.; Grishin, N. V.; Frantz, D. E.; Schneider, J. W.; Chen, S.; Li, L.; Sawaya, M. R.; Eisenberg, D.; Tycko, R.; McKnight, S.

L. Cell-free formation of RNA granules: low complexity sequence domains from dynamic fibers within hydrogels. *Cell* **2012,** *149,* 753-767.

(69) Frey, S.; Richter, R.P.; Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **2006,** *314,* 815-817.

(70) Frey, S.; Görlich, DA saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell* **2007,** *130,* 512-523.

(71) Alteri, C. J.; Xicohténcatl-Cortes, J.; Hess, S.; Caballero-Olín, G., Girón, J. A.; Friedman, R. L. *Mycobacterium tuberculosis* produces pili during human infection. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 5145-5150.

(72) DePas, W. H.; Chapman, M. R. Microbial manipulation of the amyloid fold. *Res. Microbiol.* **2012,** *163,* 592-606.

(73) Wickner, R. B.; Edskes, H. K.; Bateman, D. A.; Kelly, A. C.; Gorkovskiy, A.; Dayani, Y.; Zhou, A. Amyloids and yeast prion biology. *Biochemistry* **2012,** *52,* 1514-1527.

(74) Iconomidou, V. A.; Chryssikos, G. D.; Gionis, V.; Galanis, A. S.; Cordopatis, P.; Hoenger, A.M Hamodrakas, S. J. Amyloid fibril formation propensity is inherent into the hexapeptide tandemly repeating sequence of the central domain of silkmouth chorion proteins of the A-family. *J. Struct. Biol.* **2006,** *156,* 480-488.

(75) Parker, K. D.; Rudall, K. M. The silk of the egg-stalk of the green lace-wing fly: structure of the silk of *Chrysopa* egg-stalks. *Nature* **1957,** *179,* 905-906.

(76) Whelly, S.; Johnson, S.; Powell, J.; Borchardt, C.; Hastert, M. C.; Cornwall, G. A. Nonpathological extracellular amyloid is present during normal epididymal sperm maturation. *PLoS ONE* **2012,** *7,* e36394.

(77) Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **2009,** *325,* 328-332.

(78) Pham, C. L. L.; Kwan, A. H.; Sunde, M. Functional amyloid: widespread in nature, diverse in purpose. *Essays Biochem.* **2014,** *56,* 207-219.

(79) Chernoff, Y. O. Amyloidogenic domains, prions and structural inheritance: rudiments of early life or recent acquisition? *Curr. Opin. Chem. Biol.* **2004,** *8,* 665-671.

(80) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross-beta spine of amyloid-like fibrils.

Nature **2005,** *435,* 773–778.

(81) Jabet, C.; Sprague, E. R.; VanDemark, A. P.; Wolberger, C. Characterization of the N-terminal domain of the yeast transcriptional repressor Tup1. Proposal for an association model of the repressor complex Tup1 x Ssn6. *J. Biol. Chem.* **2000,** *275*, 9011–9018.

(82) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T.; Madsen, A.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453–457.

(83) Buchanan, L. E.; Carr, J. K.; Fluitt, A. M.; Hoganson, A. J.; Moran, S. D.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111,* 5796-5801.

(84) Mompeán, M.; Hervás, R.; Xu, Y.; Tran, T. H.; Guarnaccia, C.; Buratti, E.; Baralle, F.; Tong, L.; Carrión-Vázquez, M.; McDermott, A. E.; Laurents, D. V. Structural evidence of amyloid fibril formation in the putative agggregation domain of TDP-43. *J. Phys. Chem. Lett.* **2015**, *6,* 2608-2615.

(85) Qiang, W.; Yau, W.; Luo, Y.; Mattson, M. P.; Tycko, R. Antiparallel beta-sheet architecture in Iowa-mutant beta-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109,* 4443-4448.

(86) Petkova, A. T.; Yau, W; Tycko, RExperimental constraints on quaternary structure in Alzheimer's β-amyloid fibrils. *Biochemistry* **2006**, *17*, 498-512.

(87) Schütz, A. K.; Vagt, T.; Huber, M.; Ovchinnikova, O. Y.; Cadalbert, R.; Wall, J.; Güntert, P.; Böckmann, A.; Glockshuber, R.; Meier, B. H. Atomic-resolution three-dimensional structure of amyloid β fibrils bearing the Osaka mutation. *Angew. Chem. Int. Ed. Engl.* **2015**, *54,* 331-335.

(88) Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y. Aβ(1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **2015,** *22,* 499-505.

(89) Lasagna-Reeves, C. A.; Glabe, C. A.; Kayed, R. Amyloid-b anular protofibrils evade fibrillar fate in Alzheimer disease brain. *J. Biol. Chem.* **2011**, *286*, 22122-22130.

(90) Eskici, G.; Axelsen, P. H. Copper and oxidative stress in the pathogenesis of Alzheimer's disease. *Biochemistry* **2012**, *51*, 6289-6311.

(91) Cunningham, C. Microglia and neurodegeneration: the role of systemic inflammation. *Glia* **2013**, *61*, 71-90.

(92) Hu, K. N.; McGlinchey, R. P.; Wickner, R. B., Tycko, R. (2011) Segmental Polymorphism in a Functional Amyloid. *Biophys. J.* **2011,** *101,* 2242-2250.

(93) Lu, J. X.; Qiang ,W.; Yau, W. M.; Schwieters, C. D.; Meredith, S. C.; Tycko, R. Molecular structure of b-amyloid Fibrils in Alzheimer's disease brain tissue. *Cell* **2013**, *154*, 1257-1268.

(94) Ayala, Y. M., Pantano, S., D'Ambrogio, A., Buratti, E., Brindisi, A., Marchetti, C., Romano, M., and Baralle, F. E. Human Drosophila, and C. Elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **2005,** *348,* 575-588.

(95) Wang, H. Y.; Wang, I. F.; Bose, J.; Shen, C. K. Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics* **2004**, *83,* 130-139.

(96) Buratti, E.; Dörk, T.; Zuccato, E.; Pagani, F.; Romano, M.; Baralle, F. E. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* **2001,** *20,* 1774-1784.

(97) Passoni, M.; De Conti, L.; Baralle, M.; Buratti, E. UG repeats/TDP-43 interactions near 5' splice sites exert unpredictable effects on splicing modulation. *J. Mol. Biol.* **2011**, *415,* 46-60.

(98) Buratti, E.; Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **2001**, *276*, 36337-36343.

(99) Ayala, Y. M.; Zago, P.; D'Ambrogio, A.; Xu, Y. F.; Petrucelli, L.; Buratti, E.; Baralle FE. Structural determinants of the cellular localization and shuttling of TDP-43. *Journal of Cell Science* **2008,** *121*, 3778-3785.

(100) Winton, M. J.; Igaz, L. M.; Wong, M. M.; Kwong, L. K.; Trojanowski, J. Q.; Lee, V. M. Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *J. Biol. Chem.* **2008,** *283,* 13302-13309.

(101) Mompeán, M.; Romano, V.; Pantoja-Uceda, D.; Stuani, C.; Baralle, F. E.; Buratti, E.; Laurents, D. V. High resolution structural, stability and dynamics elucidation of the N-terminal domain of TDP-43 reveals a stable, distinct fold. *Submitted.*

(102) Buratti, E.; Baralle, F. E. Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. *Front. Biosci.* **2008**, *13*, 867-878.

(103) Cushman, M.; Johnson, B. S.; King, O. D.; Gitler, A. D.; Shorter, J. Prion-like disorders. Blurring the divide between transmissibility and infectivity. *J. Cell Sci.* **2010**, *123,* 1191-1201.

(104) Udan, M.; Baloh, R. HImplications of the prion-related Q/N domains in TDP-43 and FUS. *Prion* **2011**, *5,* 1-5.

(105) Kwiatkowski, T. J. Jr; Bosco, D. A.; Leclerc, A. L.; Tamrazian, E.; Vanderburg, C. R.; Russ, C.; Davis, A.; Gilchrist, J.; Kasarskis, E. J.; Munsat, T.; Valdmanis, P.; Rouleau, G. A.; Hosler, B. A.; Cortelli, P.; de Jong, P. J.; Yoshinaga, Y.; Haines, J. L.; Pericak-Vance, M. A.; Yan, J.; Ticozzi, N.; Siddique, T.; McKenna-Yasek, D.; Sapp, P. C.; Horvitz, H. R.; Landers, J. E.; Brown, R. H. Jr. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **2009,** *323*, 1205–1208.

(106) Deng, H. X.; Hentati, A.; Tainer, J. A.; Iqbal, Z.; Cayabyab, A.; Hung, W. Y.; Getzoff, E. D.; Hu, P.; Herzfeldt, B.M Roos, R. P.; *et al*. Amyotrophic lateral sclerosis and structural defects in Cu, Zn superoxide dismutase. *Science* **1993,** *261*, 1047–1051.

(107) Baker, M.; Mackenzie, I. R.; Pickering-Brown, S. M.; Gass, J.; Rademakers, R.; Lindholm, C.; Snowden, J; *et al*. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **2006***, 442,* 916–919.

(108) Cruts, M.; Gijselinck, I.; van der Zee, J.; Engelborghs, S.; Wils, H.; Pirici, D.; Rademakers, R.; Vandeberghe, R.; *et al*. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **2006***, 442,* 920–924.

(109) Maruyama, H.; Morino, H.; Ito,H.; Izumi, Y.; Kato, H.; Watanabe, Y.; Kinoshita, Y.; Kamada, M.; Nodera, H.; Suzuki, H.; Komure, O.; *et al*. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **2010***, 465*, 223–226.

(110) Elden, A. C.; Kim, H. J.; Hart, M. P.; Chen-Plotkin, A. S.; Johnson, B. S.; Fang, X.; Armakola, M.; Geser, F.M.; *et al*. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **2010***, 466*, 1069–1075.

(111) Ling, S. C.; Albuquerque, C. P.; Han, J. S.; Lagier-Tourenne, C.; Tokunaga, S.; Zhou, H.; Cleveland, D. W. ALS-associated mutations in TDP-43 increase its stability and promote complexes with FUS/ TLS. *Proc. Natl Acad. Sci. U. S. A.* **2010,** *107*, 13318–13323.

(112) Barnada, S. J.; Finkbeiner, S. Pathogenic TARDBP mutations in amyotrophic lateral sclerosis and frontotemporal dementia. Disease-associated pathways. *Rev. Neurosci.* **2010**, *21*, 251-272.

(113) Pesiridis, G. S.; Lee, V. M.; Trojanowski, J. Q. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **2009**, *18*, R156-R162.

(114) Zhang, Y. J.; Xu, Y. F.; Cook, C.; Gendron, T. F.; Roettges, P.; Link, C. D.; Lin, W. L.; Tong, J.; Castanedes-Casey, M.; Ash, P.; Gass, J.; Rangachari, V.; Buratti, E., Baralle, F.; Golde, T. E.; Dickson, D. W.; Petrucelli, L. Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7607-7612.

(115) D'Ambrogio, A.; Buratti, E.; Stuani, C.; Guarnaccia, C.; Romano, M.; Ayala, Y. M.; Baralle, F. E. Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res.* **2009**, *37*, 4116– 4126.

(116) Budini, M.; Buratti, E.; Stuani, C.; Guarnaccia, C.; Romano, V.; De Conti, L.; Baralle, F. E. Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich región. *J. Biol. Chem.* **2012,** *287,* 7512-7525.

(117) Romano, V.; Quadri, Z.; Baralle, F. E.; Buratti, E. The structural integrity of TDP-43 N-terminus is required for efficient aggregate entrapment and consequent loss of protein function. *Prion* **2015**, *9*, 1-9.

(118) Budini, M.; Romano, V.; Quadri, Z.; Buratti, E.; Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C- terminal Q/N prion-like domain. *Hum. Mol. Genet.* **2015,** *24,* 9-20.

(119) Zhang, Y. J.; Caulfield, T. Xu, Y. F.; Gendron, T. F.; Hubbard, J.; Stetler, C.; Sasaguri, H.; Whitelaw, E. C.; Cai, S.; Lee, W. C.; Petrucelli, L. The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet.* **2013**, *22*, 3112-22.

(120) Mackenzie, I. R.; Rademakers, R.; Neumann, M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol.* **2010,** 9, 995–1007.

(121) Geser, F.; Lee, V. M.; Trojanowski, J. Q. Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of TDP-43 proteinopathies. *Neuropathology* **2010,** *30,* 103–112.

(122) Qin, H.; Liang-Zhong, L.; Wei, Y.; Song, J. TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc. Natl. Acad. Sci.* **2014**, *111*, 18619-18624.

(123) Balbirnie, M.; Grothe, R.; Eisenberg, D. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2375-2380.

(124) Reddy, G.; Straub, J. E.; Thirumalai, D. Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 21459-21464.

(125) Thirumalai, D.; Reddy, G.; Straub, J. E. Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.* **2012**, *45*, 83-92.

(126) Yang, Z.; Shi, B.; Lu H.; Xiu, P.; Zhou R. Dewetting transitions in the self-assembly of two amyloidogenic β-sheets and the importance of matching surfaces. *J. Phys. Chem. B* **2011**, *115*, 11137-11144.

(127) Jiang, L. L.; Che, M. X.; Zhao, J.; Zhoug, C. J.; Xie, M. Y.; Li, H. Y.; He, J. H.; Hu, H. Y. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* **2013**, *288*, 19614-19624.

(128) Chen, A. K. H.; Lin, R. Y. Y.; Hsieh, E. Z. J.; Tu, P. H.; Chen, R. P.; Liao, T. Y.; Chen, W.; Wang, C. H.; Huang, J. J. Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* **2010,** *132,* 1186-1187.

(129) Capitini, C.; Conti, S.; Perni, M.; Guidi, F.; Cascella, R.; De Poli, A.; Penco, A.; Relini, A.; Cecchi, C.; Chiti, F. TDP-43 inclusion bodies formed in bacteria are structurally amorphous, non-amyloid and inherently toxic to neuroblastoma cells. *PLoS One* **2014**, *9,* e86720.

(130) Chang, C. K.; Wu, T. H.; Wu, C. Y.; Chiang, M. H.; Toh, E. K. W.; Hsu, Y. C.; Lin, K. F.; Liao, Y. H.; Huang, T. H.; Huang, J. J. T. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem. Biophys. Res. Comm.* **2012,** *425*, 219-224.

(131) Hida, T.; Nakamura, F.; Usui, H.; Takeuchi, K.; Yamashita, N.; Goshima, Y. Semaphorin3A-induced axonal transport mediated through phosphorylation of Axin-1 by GSK3b. *Brain Res.* **2015,** *1598,* 46-56.

(132) Lukavsky, P. J.; Daujotyte, D.; Tollervey, J. R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F. E.; Damberger, F. F.; Allain, F. H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43 *Nat. Struct. Mol. Biol.* **2013**, *12,* 1443-1449.

Chapter 2

Combining Classical MD and QM Calculations to Elucidate Complex System Nucleation: A Twisted, Three Stranded, Parallel β-sheet Seeds Amyloid Fibril Conception

nani gigantum
humeris insidentes.

2.1 Abstract

The crystal structure of the Sup35 prion segment, GNNQQNY, revealed precise side chain packing and an extensive H-bond network. However, the conformers and stabilizing interactions involved at nascent amyloid formation are still unclear. Here, long molecular dynamics simulations and quantum mechanical calculations have been utilized to study the conformation and energetics of the initial structure that acts to nucleate further growth. Considering all the plausible intermediates that may act as stepping-stones, we find that the initial nucleus is a twisted single-layer, three-stranded parallel β-sheet. H-bonds between β-strands in this twisted sheet, some of which differ from those of the crystal structure's non-twisted β-strands, are key for the nucleus' formation and stability. High level theoretical calculations of these H-bonds' energetics can account for this amyloid-like trimer's remarkable stability. The intermeshing of facing sheets to form the dry interface provides less stability and would occur between two three-stranded β-sheets without meta-stable water nanowires.

2.2 Introduction

Amyloids are implicated in a score of human illnesses including Alzheimer's and Parkinson's diseases. In the last decade, atomic structures of some amyloid fibrils have been determined by ssNMR or crystallography[1,2], yet the pathway and energetics for forming the structural kernel which seeds further amyloid growth are unresolved. Many amyloids such as Aβ implicated in Alzheimer's disease are predominately hydrophobic, but others like GNNQQNY are highly polar which makes this segment from the yeast "prion" Sup35 an interesting puzzle for understanding the physical basis of the formation and stability of complex systems.

Based on the exceptionally high packing efficiency of its side chains at the dry interface between two β-sheets, Eisenberg and coworkers proposed that this side chain intermeshing is crucial for amyloid formation and stability[3,4]. Using ballpark estimates of enthalpy and entropy changes, Nelson *et al.* [3] speculatively proposed that the Sup 35 amyloid nucleus would consist of two β-sheets, each composed of two β-strands, whose side chains pack at the dry interface. Somewhat later, based on results obtained using the straight β-sheets of the crystal structure and DFT calculations with plane waves, the main driving force for GNNQQNY amyloidogenesis was proposed to be the cooperative polarization and hyper-stability of the intermolecular H-bonds[5]. These results provide the basis for an intuitively attractive model for amyloid formation based on the slow formation of a structural nucleus and then a faster growth step. However, more recent investigations reveal that the GNNQQNY amyloid fibril is moderately twisted in solution and that the experimentally observed straightness of the β-sheets is a consequence of crystal packing forces[6]. The effects of terminal charges, ionic strength, strand configuration, mutation and temperature on GNNQQNY oligomerization, fibrillization and crystallization have already been investigated in a score of publications using high resolution experimental[7-9] and computational[10-11] methods (see **Sup. Table 1** for a more complete survey). In addition, studies of GNNQQNY association to form a dimer[12] and to elongate the fibril[13] have provided some details on the nucleation process. Nevertheless, some key questions involving the nature of the first stable intermediate key for oligomer formation, the expulsion of water as the "dry" interface forms, and the possible energy bonus for stabilizing H-bonds in the growing oligomer are still open. These questions are addressed here using a novel combination of independent, time-extensive molecular dynamics simulations and quantum mechanical (ONIOM[14] (DFT:PM6)) calculations to evaluate the structural stability of the key putative structural intermediates, using B3LYP/6-311+G(d,p) for the QM partition. Here, we show that: (i) a twisted, three stranded, parallel β-sheet is the key intermediate for amyloidogenesis, (ii) straight versus twisted β-sheets have different hydrogen bonding networks and the latter are significantly more stable and (iii) water molecules readily escape as two three-stranded β-sheet intermediates come together to form the amyloid fibril nucleus, without the formation of trapped meta-stable water nanowires[15].

2.3 Results and Discussion

There are two fundamental ways that two Sup35 peptides may combine at the start of amyloid formation. They could stack to form a parallel two-stranded β-sheet. This "sheet" dimer, which we call **"S2", as is composed of two strands,** is stabilized by multiple side chain and backbone H-bonds.

Alternatively, the two strands could associate laterally to permit the intermeshing of side chains at what will become the "dry interface" in the mature amyloid fibril. We refer to this dimer as **"S1·S1", where "·" denotes interaction through sidechain intermeshing.** Tight side chain packing stabilizes **S1·S1**. We investigated the stability of these dimers by subjecting them to MD simulations. The **S1·S1** dimer dissociates quickly, as evidenced by a large increase in RMSD after 10 ns or less of simulation time [**Sup. Fig. 1 & 2**]. Thus, it appears that the stability provided by tight side chain packing is insufficient to hold the two peptides together long enough for further growth. In contrast, the **S2** dimer is more stable; dissociation was never seen before 65 ns [**Sup. Fig. 3**]. A consideration of the bimolecular association kinetics (detailed in the Sup. Material) provides evidence that when the GNNQQNY concentration is above 3.5 mM, 65 ns would be long enough for a free monomer to encounter another monomer or **S2** dimer. These simulations have been repeated with different initial velocities and conformations, and similar results were always observed [**Sup. Fig. 1 & 2**].

The stacking of an additional peptide on the **S2** dimer could produce a three stranded, all parallel β-sheet. We call this trimer **S3**. When subjected to MD, **S3** showed a remarkable stability, with no dissociation occurring even after 500 ns. This result is curious considering that parallel β-sheets, due to their suboptimal N-H | | | O geometry, are usually thought to be less stable than antiparallel β-sheets. An examination of the structures formed along the simulation revealed the source of **S3'**s remarkable stability. After few ps, during the equilibration period, the sheet *twists* significantly, which improves the geometry of the H-bonds and changes the pairings of some side chain H-bonds.

In the twisted structure of ***S3,*** which we write in *italics* to emphasize its twist, the sidechains of the central strand anchor the structure by forming H-bonds with residues of both edge strands (**Fig. 2**, **top**). Once it twists, ***S3*** remains structurally stable, that is, no significant conformation changes are seen for the remainder of the 500 ns simulation time. In the hexamer consisting of two ***S3*** trimers packed in an antiparallel fashion, the sidechain packing between the two sheets is still remarkably tight (the surface complementarity, Sc, is 0.83)[6], almost as tight as the hexamer formed by two non-twisted **S3** trimers (Sc=0.86)[3] and the percent solvent accessible surface area is only slightly higher in ***S3*** (56%) than **S3** (53%).

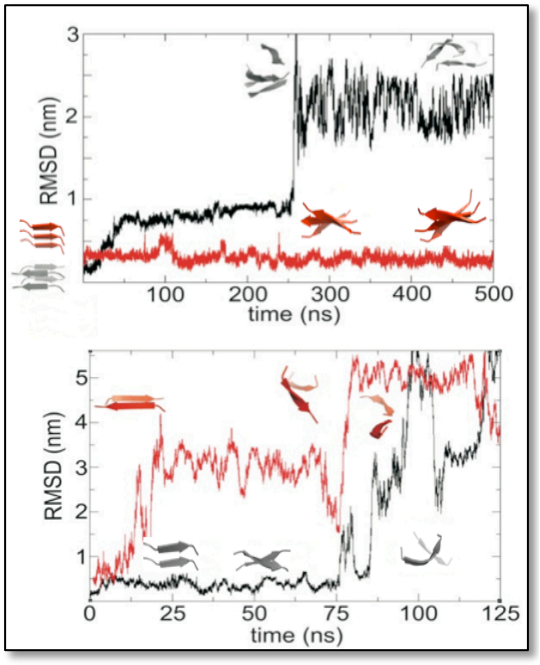


Figure 1. T*op*: RMSD showing the structural stability of ***S3*** and **S2·S2** systems for a 500 ns run. For the **S2·S2** system the most representative trajectory from five simulations performed is depicted, in which one strand dissociates and the other three seem to evolve towards an **S3**-like system. For the ***S3*** system, six simulations were performed and all the results showed that this timer is the remarkably stable. The structures are shown at time 0, 250 and 500 ns. ***Bottom*:** RMSD for **S2** and **S1·S1** showing selected trajectories. From six runs on each system, the **S2** simulation with the lowest kinetic stability (*i.e.* quickest dissociation) for **S2** and the **S1·S1** simulation with the highest kinetic stability (*i.e.* slowest dissociation) **are shown**, in order to emphasize that the intermeshing only does not provide with enough stability (see *Sup. Info.* for further discussion). The structures shown correspond to snapshots extracted at 0, 50 and 100 ns.

We have simulated the behavior of a trimer called **S2·S1**, with the **S2** dimer combined laterally with an additional monomer. This **S2·S1** trimer is structurally unstable, highlighting the necessity of larger systems to form a stable dry interface (Sup. Fig. 5).

We have also considered a tetramer called **S2·S2**, which can be regarded either two **S2** dimers combined laterally through the so-called dry interface, or two **S1·S1** dimers stacked and stabilized by H-bond formation. This **S2·S2** system is structurally unstable and evolves towards conformations that resemble ***S3***, or other single beta-sheet conformers. This indicates that the side chain packing of four monomers is insufficient to form a stable dry interface. **Figure 1** shows the RMSD for the most representative trajectories of **S2**, **S1·S1**, ***S3*** & **S2·S2**, whereas **Sup. Fig. 3, 4 and 5** contains the set of repetitions to assess the reproducibility of the results.

The hydrogen bonding pattern that is present in the crystal structure changes upon twisting, as shown in **Figure 2**. We have determined that these changes make a strongly stabilize ***S3*** by -8.92 kcal/mol relative to S3 (**Fig. 2**).

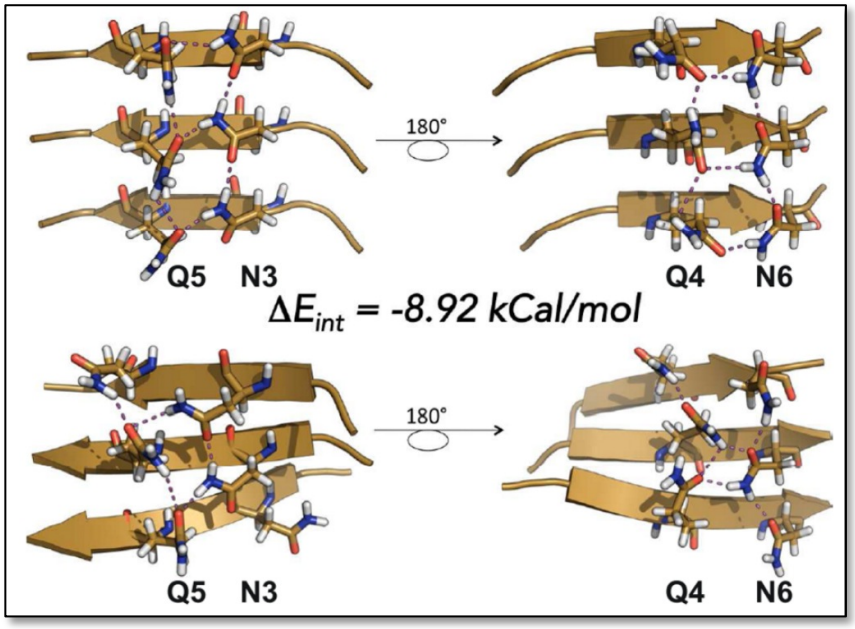


Figure 2. H-bonding pattern for the **S3** structure, resembling the experimental data (top); the pattern for the ***S3*** twisted system revealed by the simulations (bottom), each viewed from the perspective of the dry (left) and wet (right) interfaces, respectively. The gain in interaction energy due to twisting is -8.92 kcal/mol, consistent with more compact sheets due to change in the HB pairings–see *Methods* in SI. –

Based on its elevated stability, we conclude that the twisted ***S3*** trimer is the key intermediate for forming the minimal seed of the Sup35 amyloid fibril, which we propose to be a hexamer of two ***S3*** trimers packed in an antiparallel manner [**Fig. 3, bottom**] In ***S3***, residues Q4 and N6 show low mobility, as gauged by their small RMS fluctuations [**Sup. Fig. 6**]. We propose that this low dynamic motion of these residues predisposes them to form van der Waals contacts when two trimers come together, thus facilitating the formation of the dry interface of the hexamer ***S3·S3*.** Moreover, the point substitution of both the central Asn or Gln or both by Gly or Tyr prevents aggregation, as determined theoretically [**Sup. Fig. 7**] and experimentally using NMR [**Sup. Fig. 8**]. The NMR characterization of GNNQQNY under solution conditions similar to those employed for crystallography studies demonstrates that the peptide adopts a broad structural ensemble which is dominated by disordered conformers [**Sup. Fig. 8**]. This is consistent with the behavior of GNNQQNY monomers observed by MD.

In the yeast prion Sup35, the amyloidogenic sequence GNNQQNY occurs as a segment within a large protein containing 685 residues[16]. Considering this, we have performed the simulations of GNNQQNY described above with the N- and C-termini neutralized by acetyl and amide groups, respectively (*i.e.* **CH_3-CO-**HNGly1…Tyr7-CO-**NH_2**). However, since the crystal structure of GNNQQNY was solved with charged termini, and as this form is often studied experimentally, we have carried out additional simulations to study how charged termini (*i.e.* $^{+}NH_3$-Gly1…Tyr7-COO^{-}) affect the formation of its smallest oligomers. In the dimer consisting of a layer of peptides arranged laterally in an anti-parallel manner to permit side chain packing (**+S1·S1-**), the

proximity of oppositely charged groups stabilizes the structure [**Sup. Fig. 9**]. It resists dissociation and significant structural perturbation, in contrast to the behavior of the uncharged dimer **S1·S1** [**Fig. 1**].

Interestingly, the **+S2-** dimer, despite the close proximity of charged groups of the same sign, is still stable, as was seen in previous studies[17]. Thus, charged termini stabilize the L2 dimer but do not decisively destabilize the S2 dimer. A discussion on the effect of the electrostatic interactions in **S2**-like system is given in the **Sup. Info**.

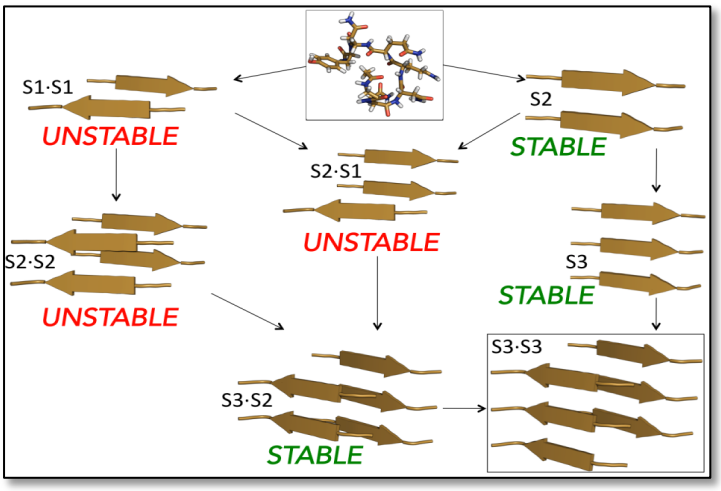


Figure 3. Intermediates considered in the proposed pathway. Systems have been labelled as "stable" or "unstable" based on our simulations.

The direct characterization of the slow formation of a seed for oligomerization and amyloidogenesis, which requires unaffordable computational times, has been tackled here by considering cross-β spine-like structures and testing their stabilities. The instability of **S1·S1**; **S2·S1** and **S2·S2** arises from the low energetic contribution from dry interface formation. This will usually occur after ***S3*** forms, as ***S3***'s side chains are correctly predisposed for packing, and the interaction energy for two ***S3*** trimers forming the dry interface is highly favorable, as determined by QM methods (-30.46 kcal/mol). This result, and ***S3***'s lack of dissociation during long MD runs, allowed us to propose that ***S3*** is the structural species seeding amyloid conception in GNNQQNY. Whereas a ***S3*** trimer and a **S2** dimer might combine to form an ***S3·S2*** pentamer with a marginally stable dry interface (ΔE_{int} = -8.42 kcal/mol), [**Sup. Fig. 10** and **Sup. Table 2**], two ***S3*** trimers could combine to form an ***S3·S3*** hexamer with proper sidechain packing and a much more stable dry interface (ΔE_{int} = -30.96 kcal/mol), [**Sup. Fig. 11** and **Sup. Table 2**].A 100 ns simulation of four of these twisted trimers naturally evolved towards the intermeshed nucleus, in absence of biasing potentials, as shown in **Fig. 4**. The trimers maintain their stable conformation, as depicted in **Sup. Fig. 11 & 12**, and form the dry interface. Once formed, the ***S3·S3*** hexamer is stable, as evidenced by low RMSD values observed during the course of a 100 ns simulation (**Sup. Fig. 13**). Previous studies

suggested that meta stable water molecules may be trapped between two long GNNQQNY β-sheets (made of eight β-strands or more each) as they come together to form the dry interface[15]. These water nanowires persist for *ca.* 20 ns and retard the formation of the dry interface. Our results strongly suggest that dry interface formation can readily occur between two three stranded β-sheets. Since water nanowires are much less stable between shorter sheets, water molecules are quickly expelled and the dry interface can form much more quickly. Based on these considerations, we conclude that the dry interface formation generally occurs between two β-sheets as small as trimers rather than between long β-sheets.

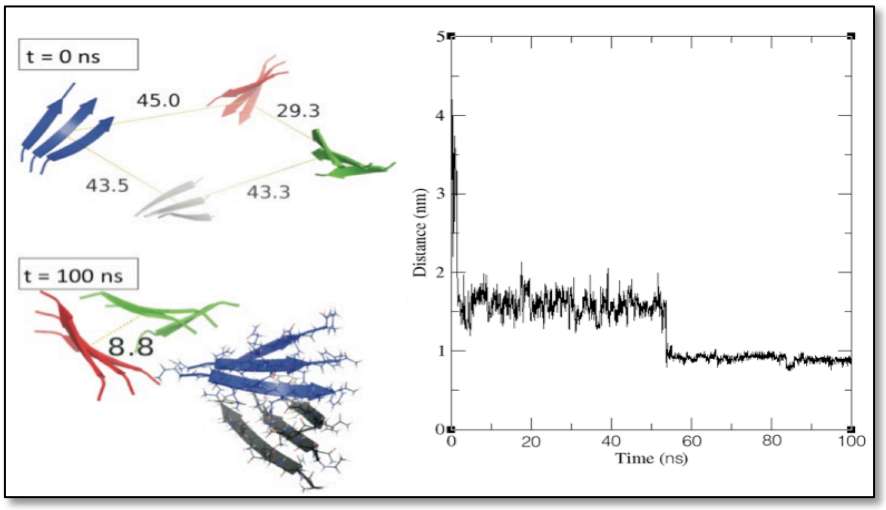


Figure 4. Formation of the dry interface. Four ***S3*** β-sheets were placed far apart in a simulation box (top left panel). During the course of the simulation, the ***S3*** β-sheets moved closer together as shown by the decreasing distance separating the Q4 Cα atoms in the middle β-strand of two different ***S3*** trimers (right panel). By 100 ns, the side chains had interdigitated to yield two ***S3·S3*** whose central Q4 Cαs are separated by ~8.8Å, which corresponds to the distance (represented by red and green arrows in the lower left panel) expected for the nucleus that serves for further growth. The other ***S3·S3*** system (black & blue) is shown as ribbons (backbone), and as sticks (side chains) to highlight the side chain intermeshing. This ***S3·S3*** system resembles the minimal cross-β spine that can form and it is structurally stable over the course of a 100 ns MD run (see **Sup. Fig. 13**).

2.4 Methods

NMR Spectroscopy: In summary, 1D ^{1}H and 2D TOCSY, ROESY, ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC NMR spectra were obtained on the peptides GNNQQNY and GQNNGQNY (negative control) using 600 and 800 MHz Bruker spectrometers equipped with cryoprobes. Both peptides were purchased from Genescript Inc. 1D ^{1}H spectra were used to follow aggregation. After assignment, the ^{1}H, ^{13}C and ^{15}N chemical shifts were analyzed for tendencies for partial structure formation using the TALOS+ package[18].

MD Simulations: Briefly, the GROMACS package (ver. 4.5.5.)[19] using the parameters of the amber99sb-ildn force field[20] were used to perform MD calculations. The peptide models for various GNNQQNY peptide oligomers, based on the 1YJG PDB structure[3], were correctly solvated and pre-equilibrated at 1 ATM of pressure and 300 K prior to the main 100 to 500 ns MD runs.

ONIOM Calculations: In a nutshell, the ONIOM approach[14], which divides the systems into regions of low and high theory, was used to study the energetics of GNNQQNY peptide oligomers. High level DFT/(B3LYP/6-311+G(d,p)) was utilized for the side chain amide atoms and the PM6 semi-empirical method was used for all other atoms. These calculations were carried out the solvation continuum model (PCM) with the Gaussian09 package[21]. Due to their high computation cost, these calculations were applied to the most interesting oligomers; namely *S3*, *S2·S3* and *S3·S3*. The solvent accessible surface area was calculated using the Lee and Richard's algorithm[22] as implemented in the MOLMOL computer program using a solvent probe radius of 1.4 Å.

For more details on the experimental and computational procedures, please see the Supporting Information section.

2.5 Acknowledgments

This work is funded by grant CTQ 2010-21567-C02-02 from the Spanish Ministry of Science & Innovation.

2.6 References

(1) Tycko, R. Solid-state NMR studies of Amyloid Fibril Structure *Annu. Rev. Phys. Chem.* **2011**, *62*, 279-299.

(2) Eisenberg, D.; Junker. M. (2012) The amyloid state of proteins in human diseases. *Cell* **2012**, *148*, 1188-1203.

(3) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A.O.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **2005**, *435*, 773-778.

(4) Sawaya, M.R.; Sambashivan, S.; Nelson, R.; Ivanova, M.I.; Sievers, S. A.; Apostol, M.I.; Thompson, M.J.; Balbirnie, M.; Wiltzius, J.J.; McFarlane, H. T.; Madsen, A.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007**, *447*, 453-457.

(5) Tsemekhman, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* **2007**, *116*, 761-764.

(6) Esposito, L.; Pedone, C.; Vitagliano, L. (2006) Molecular dynamics analyses of cross β-spine steric zipper models: β-sheet twisting and aggregation. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 11533.

(7) van der Wel, P. C. A., Lewandowshi, J. R.; Griffin, R. G. Solid state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J. Am. Chem. Soc.* **2007,** *129*, 5117-5130.

(8) Marshall, K. E.; Hicks, M. R.; Williams, T. L.; Hoffmann S. V.; Rodger, A.; Dafforn, T.; Serpell, L. Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: Structural changes accompany a fiber-to-crystal switch. *Biophys. J.* **2010**, *39*, 330-338.

(9) Portillo, A. M.; Krasnoslobodtsev, A. V.; Lyubchenko, Y. L. Effect of electrostatics on aggregation of prion protein Sup35 peptide *J. Phys. Condens. Matter* **2012** *24,* 165205.

(10) Zheng, J.; Ma, B.; Tsai, C.; Nussinov, R. Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from yeast prion Sup35 *Biophys. J.* **2006**, *91*: 824-33.

(11) Zhang, Z.; Chen, H.; Bai, H.; Lai, L. Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35 *Biophys. J.* **2007,** 93, 1484-92.

(12) Reddy, A.S.; Chopra, M.; dePablo, J.J. GNNQQNY Investigation of Early Steps During Amyloid Formation. *Biophys. J.* **2010**, 98, 1038-45.

(13) Reddy, G.; Straub, J.E.; Thirumalai, D. (2009) Dynamics of Locking of Peptides onto Growing Amyloid Fibrils. *Proc. Natl. Acad. Sci.* **2009**, 106: 11948-11953.

(14) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: TEOCHEM* **1999**, *461-462*: 1.

(15) Reddy, G.; Straub, J. E.; Thirumalai, D. Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl. Acad. Sci. USA* **2010**, *107,* 21459-21464.
(16) Balbirnie, M.; Grothe, R.; Eisenberg, D. (2001) An amyloid forming peptide from the yeast prion Sup35 reveals a dehydrated β-sheet structure for amyloid. *Proc. Natl. Acad. Sci. USA* **2001**, 98, 2375-2380.
(17) Vitagliano L, Esposito L, Pedone C, De Simone A. (2008) Stability of single sheet GNNQQNY aggregates analyzed by replica exchange molecular dynamics: antiparallel versus parallel association *Biochem. Biophys. Res. Commun.***2008***, 377,* 1036-1041.
(18) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. (2009) TALOS+: A hybrid method for predicting protein backbone torsión angles from NMR chemical shifts. *J. Biomol. NMR.* **2009**, *44*, 213-223.
(19) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(20) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Kepleis, J.L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950.
(21) Frisch, M. J., Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. *et al.* Gaussian 09 Revision A.2, **2009**, Gaussian Inc. Wallingford CT. USA.
(22) Lee, B; Richards, FM. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.,* **1971,** *55*, 379-400**.**

2.7 Supporting Information

2.7.1 Supporting Methods

NMR Spectroscopic Characterization of GNNQQNY and GQNNGQNY peptides

To characterize the initial conformation of GNNQQNY and to corroborate the prediction based on our MD results that the rigidity and interactions of the central residues play key roles in nascent GNNQQNY oligomerization, we studied the conformation of this heptapeptide and a negative control peptide, in which the position of Gly1 and Gln4 are switched, by NMR spectroscopy. By placing Gly, a flexible residue whose side chain is unable to replace the interactions formed by Gln's side chain, we expect that oligomer formation will be blocked. To prevent the formation of pyroglutamate at the N-terminus, we added an additional Gly at the N-terminus. Thus the sequence of this negative control peptide is Gly0-**Gln1**-Asn2-Asn3-**Gly4**-Gln5-Asn6-Tyr7, with the switched residues written in **bold**. This peptide and the Sup35 wild type heptapeptide: Gly1-Asn2-Asn3-Gln4-Gln5-Asn6-Tyr7, were obtained from Genescript Corp. They are over 95% pure (HPLC) and their identities were confirmed by MALDI-TOF mass spectrometry and NMR spectroscopy.

NMR spectra were acquired on a Bruker 600 & 800 MHz (^{1}H) spectrometers equipped with a triple resonance cryoprobes and Z-gradients at 25.0 ºC, in 90% H_2O/ 10% D_2O with 50 μM DSS as the internal chemical shift standard, 10 mM $ZnCl_2$ and 10 mM deuterated sodium acetate/acetic acid (pH 5.1). These solution conditions are similar to those used for crystallization by Eisenberg and coworkers[1]. The peptide concentration was 3.44 mM for the wild type peptide and slightly higher, 3.67 mM, for the negative control peptide. We acquired the following spectra: 1D ^{1}H (32 transients) and 2D ^{1}H TOCSY (mixing time 60 ms, 8 transients per increment, 512 increments in the 2nd dimension), 2D ^{1}H ROESY (spinlock time 300 ms, 64 transients per increment, 512 increments in the 2nd dimension), ^{1}H-^{13}C HSQC (256 transients per increment, 128 increments in the ^{13}C dimension) and ^{1}H-^{15}N HSQC (512 – 1024 transients per increment, 128 increments in the ^{15}N dimension).

Bimolecular Association Kinetics Calculation

An estimate for the bimolecular association rate constant can be derived from Fick's first law of diffusion:

$$\mathbf{k = 4\,\pi\, r_o\,(D_{S+}D_{S2})\cdot 10^{-3}\,N_o}$$

where r_o is the distance between the centers during the encounter (assumed to be a 1 nm), D_A is the diffusion constant (about $32.5*10^{-7}$ for a 840 Da peptide and about $25.2\cdot10^{-7}$ for the 1680 Da dimer at 20 ºC in H_2O) and No is Avogadro's number. This gives a value of $4.37*10^9$ $M^{-1}\cdot s^{-1}$ for the bimolecular association constant. From this value, it can be calculated that there would be 284 contacts per 65 ns at 1 M concentration, or 1 contact per 65 ns at 3.5 mM of heptapeptide. We think that this

estimate is a ballpark figure; all encounters are unlikely to be productive, so a higher concentration of peptide might be needed to drive oligomerization. On the other hand, four of the six simulations of the S2 dimer did not dissociate even after 100 ns. Therefore, it is likely that the monomer would, on average, have longer than 65 ns to encounter S2, and therefore, a lower monomer concentration would be required to further oligomerization.

Molecular dynamics simulations

The models for the peptide structures are based on the X-ray crystal experimental data[1] (PDB code 1YJP), and were generated with PyMOL. Since previous work has shown the importance of electrostatics in the aggregation of Sup35[2] we have capped the termini residues of each peptide in order to mimic the behaviour of the full peptide sequence. GNNQQNY is just a short heptapeptide segment in a loop of a large 680-residue protein. Thus, in order to represent more accurately the behavior of this aggregating segment, the peptide was amidated at the C-terminus and acetylated in the N-terminus. However, as the peptide with charged termini has been studied extensively by crystallography and computational methods, we have also performed additional simulations with the terminal charges present.

For all the MD simulations, we employed the GROMACS package, version 4.5.5.[3] The structures were placed in a cubic box leaving at least 1.0 nm between any atom of the solvent molecules and the box edges. Parameters from the amber99sb-ildn force field[4] were applied to all the simulated systems, solvated with TIP3P[5] water molecules. The choice of this force field is based on its appropriateness for amyloid-like systems[6]. Short-range nonbonded interactions were cut off at 1 nm, with long-range electrostatics calculated using the particle mesh Ewald (PME) algorithm[7] and application of dispersion-correction to account for van der Waals interactions at distances longer than the cut off. Periodic boundary conditions were applied in all directions.

After steepest descents minimization, two equilibration periods were done using the Berendsen[8] coupling algorithm. Once the temperature and pressure were properly set to 300 K and 1 bar (500 ps and 2 ns runs, respectively) with heavy atom positions restrained with the LINCS[9] algorithm, the MD simulations were produce with the Nosé-Hoover[10] thermostat and the Parrinello-Rahman[11] barostat in absence of restraints, with time constants of 0.5 and 1.0 ps, respectively. A time step of 2 fs was used in both the equilibration stages and the MD run. **Distinct initial velocities were utilized to ensure the independence of the different MD runs starting on the same structure. This is a standard procedure to allow the system, which starts from the same structure, to evolve through different trajectories thereby resulting in different simulations. This procedure is done at the start of a simulation; just after defining the topology of the system is defined and following the energy minimization step, one uses a random seed to generate random initial velocities that will allow the system to reach a Maxwell distribution.** The root

mean square displacement was calculated for all of the simulated systems to assess the kinetic stability of all the structures.

To further corroborate these results, we performed a series of additional simulations. Even though the 500 ns simulations are in principle long enough to consider proper sampling, it has been recently shown that in some cases performing multiple simulations for shorter periods of time is a better gauge of the reproducibility[12]. By using both approaches, we can ensure that a thorough sampling was done.

These additional runs corroborate that **S2** is clearly more stable than **S1·S1**. The latter exhibits a dual behavior: it either dissociates or rearranges into an anti-parallel β-sheet, but none of these simulations lasts longer than 10 ns as a "layer", which is the proposed biological assembly[1]. In the layer-like structure (*i.e.*, **S1·S1**) the stabilizing interactions are due to side chain intermeshing and not due to the formation H-bonds as in the S2-like antiparallel structure. The rapid dissociation or alternation of the **S1·S1** structure indicates that these side chain interactions are relatively weak. On the basis of its ephemeral stability, we conclude that the possibility of the **S1·S1** intermediate persisting long enough so that other monomers can join it is negligibly small, when it is uncharged. See below for a description of the behavior when the terminal charges are present (**Sup. Fig. 1 & 2**).

For ***S3***, since the RMSD reveals a remarkable stability in multiple 100 ns runs, all of which showed similar behavior, as well as a long 500 ns simulation, we conclude that this structural motif lasts long enough to be the most important intermediate in the formation of the seed for aggregation (**Sup. Fig. 3**).

S2·S2 showed significant structural changes during the course of the 500 ns run; one monomer dissociates and the other three evolve towards an ***S3***-like configuration. The results reported in Supplementary Figure 4 show that in only *one out of five 100 ns simulations* did the motif remain stable without any significant structural rearrangement.

To address whether the dry interface could be formed in some intermediate structures not considered in the above cases, two additional oligomers, named ***S2·S1*** and ***S3·S2*** were tested. The first consists of a **S2** dimer with a monomer docked laterally through sidechain packing, whereas the last is a **S3** plus **S2** interacting system through their intermeshed sidechains. Both the two system consider the dry interface formation.

Supporting Figure 5 describes the behavior of ***S2·S1*** through the RMSD, considering the situations in which the peptides are either charged or with the terminal residues neutralized. Similarly, Sup. Figure 6 explains the results observed of the analogous ***S3·S2*** system. For the ***S2·S1*** system, a dry interface cannot be formed, whereas in the case of an ***S3·S2*** arrangement some effective sidechain intermeshing was found to take place (Sup. Fig. 6). However, its structural stability is considerably lower than that of the ***S3·S3*** system, hence we conclude that the dry interface is most likely to definitely form when the mating sheets have a minimal length of three strands (**S3**).

We conclude that the ***S3*** trimer is to be the key intermediate leading to the ***S3·S3*** nucleus for the following reasons: 1) The dry interface is not prone to form with two or four monomers, **S1·S1** or **S2·S2**, respectively), 2) The ***S3*** trimer is stable in the long 500 ns simulation and multiple 100 ns simulations, and 3) Two *S3* trimers were observed to combine to form ***S3·S3*** (**Fig. 4** & **Sup. Fig. 11**). We consider ***S3·S3*** to be the "nucleus", since it is the smallest structure showing the amyloid-like properties derived from experimental data in a solvated context, that is, similar H-bonding pattern, amide stacking and sidechain intermeshing (*i.e.*, dry interface formation) as well as a high stability as evidenced by its lack of dissociation or structural alteration after 100 ns of simulation (**Sup. Fig. 12).**

ONIOM calculations

The ONIOM[13] methodology permits the division of the system. Here, we made a two-region partition to use two different levels of theory, resulting in the so-called low and high layers. The former was modelled with the PM6[14] semiempirical method, whereas we employed DFT for the latter. For these QM calculations, we optimized the geometry with the B3LYP functional and the 6-311+G(d,p) basis set, accounting for dispersion and polarization effects. This functional was chosen for the geometry optimization procedure since it is well known to yield very good results when applied to organic systems. Recently, a very similar approach to dissect L-asparaginase II's enzymatic mechanism has been described[15].

The high layer consists of all side chain amide groups, since they are involved in the H-bonding and experience hyperpolarization upon monomer addition[16]. Thus, a monomer contains 5 amide residues times 5 amide atoms = 25 QM atoms. Accordingly, ***S3*** contains 75 QM atoms and ***S3·S3***, 150.

Since PM6 can deal with β-sheet backbone–backbone hydrogen bonds, the low layer includes the rest of the peptide, *i.e.*, capping atoms CH_3CO- and -NH_2, first and last residues of each strand (Tyr and Gly), and amide backbone atoms.

With respect to the basis set, even though the inclusion of diffuse functions has been reported to substantially raise the computational cost without notably increasing the optimized geometries[17], we decided to perform our calculations with these functions on heavy atoms. To this end, we performed our calculations in a multi-step way to ensure a smooth convergence. First, we froze the Cα atoms to keep the backbone at its experimental position and used the 6-31+G(d) basis set. After that, we employed that optimized structure to be re-optimized with the larger basis set 6-311+G(d), adding 18 extra basis functions to the former. Finally, we performed a 6-311+G(d,p) calculation without any restraints, which adds polarization functions to hydrogen atoms with respect to the previous steps. By this procedure, we obtained the fully relaxed geometry with the appropriate description of the H-bonds between amide groups.

Having seen that the **S3** trimer has a remarkable stability upon twisting (***S3***), one may ask what is the energetic bonus afforded by this structural rearrangement. The hydrogen bonds between the strands within a sheet are stronger than those for water in ice, as determined by Baker *et al*.[16] However, in that manuscript it was advanced that the physiological twisted fibrils may possess higher interaction energy, since those calculations where performed in gas phase and with boundary conditions to account for an infinite system. Moreover, no relaxation/optimization was described, presumably due to the high computational cost. To properly quantify the effect of the twist, we followed a subtractive scheme similar to that of Martín-Pintado *et al*.[18] We first defined the **S3** trimer as State A, which reproduces the crystallographic experimental data. State B is the twisted structure resulting from averaging over the 500 ns trajectory, ***S3***. The all-atom contribution to the intra- and intermolecular interaction energies within the strands can be obtained from the fully optimized straight and twisted trimers, named States A & B, by sequential removal of the strands (**Figure 2**, main text). The overall energy difference between ***S3, S2, S1*** **& S3, S2, S1** with respect to the trimers allows for a direct estimation of the energetic contribution of the twist to the global stability (*i.e*., the gain in interaction energy due to the different hydrogen bonding pattern.)

The choice of the averaged structure as representative of the twisted conformer is based on the low RMSD obtained when fitting the frames of the trajectory to this system, as depicted in **Sup Fig.12**.

The kinetic MD-based study as well as the ONIOM calculation showed that the trimer ***S3*** is extremely stable and we propose that this is the key kernel structure for amyloidogenesis. The dry interface can only be effectively formed when two of these units mutually interact, since the energy contribution of their sidechain-sidechain contacts is too low when less than three strands per sheet are present. To quantify this value, the interaction energy of two ***S3*** trimers combining to yield the nucleus ***S3·S3*** is a direct measure of the energy required for the dry interface formation. To estimate this magnitude, we again performed ONIOM calculations, to keep all the results at the same level of theory. For the ***S3*** system, the choice of the averaged structure is justified based on the remarkably low RMSD value (**Sup. Fig. 13**). The ***S3·S3*** structure was also averaged over the last 40 ns for this purpose, and to be sure that we were computing the interaction energy with the corresponding systems in local minima, we also performed frequency calculations that corroborate the absence of transition states in both the two structures. To our knowledge, this is the first time that such time-consuming force constant calculations are performed in a system of this size.

With respect to the ***S3·S2*** system, which exhibits subtle structural distortion although a dry interface might be formed, the interaction energy for these contacts have been computed over different snapshots of the last 10 ns of the trajectory, averaging the corresponding values (See **Sup. Table 2.)**

The gain of interaction energy upon twisting was also calculated at the M06-2X/6-311+G(d,p) level, since this functional[19] was derived for the computation of non-covalent interactions and has been successfully applied to biological systems[18]. This procedure yielded very similar results (**Sup. Table 2**).

All the calculations were performed using a solvation continuum model (PCM) with the Gaussian09[20] package, allowing the effect of solute polarization to be included and avoiding the higher energies derived from gas-phase calculations.

Charged vs. Capped Systems: Two distinct pathways towards the seed formation

We have shown that the smallest stable structural motif that can act as a seed for further nucleation in Sup35 is a *twisted* three-stranded parallel β-sheet (***S3***). The kinetic study shows that the most plausible pathway towards the formation of the ***S3·S3*** seed of Sup35 can be schematized as:

Monomer (x2) → ***S2*** → ***S3*** → intermeshing (***S3·S3***)

The alternative pathway:

Monomer (x2) → **S1·S1** → ***S2·S2*** → ***S3·S3***

is shown to be kinetically unstable.

This is the situation for the peptides with neutral termini (*i.e.*, in the context of the full-length protein). However, it seems that for the heptapeptide with charged termini, both are plausible since the electrostatic interactions help hold the two chains together, stabilizing the side chain intermeshed conformation against dissociation.

To explain how the electrostatic repulsion that arises from the proximity of charged termini is overcome, we have simulated the dynamics of +GNY-, +GNNY-, +GNNQY-, +GNNQQY- & +GNNQQNY- peptides in a two-stranded parallel β-sheet. We have run 100 ns of MD on each structure, with the same settings that were listed in the Molecular Dynamics Simulations section. Taking into account that 11 H-bonds are formed between two strands (6 backbone-backbone & 5 sidechain-sidechain), we could determine the minimal number of H-bonds necessary to overcome the electrostatic repulsion. As summarized in **Table S2** at least 7 H-bonds (arising from 2 GNNQY pentapeptides) are needed to achieve stability.

Considering growing β-sheets composed of charged +GNNQQNY- peptides, they can be expected to be more stable due to hyperpolarization of the H-bonding amide groups. Twisting, which manifests itself in dimers, trimers and larger oligomers, will also stabilize β-sheets composed of peptides with charged termini since twisting separates groups carrying like charges, thus reducing their electrostatic repulsion.

2.7.2 Supporting Figures

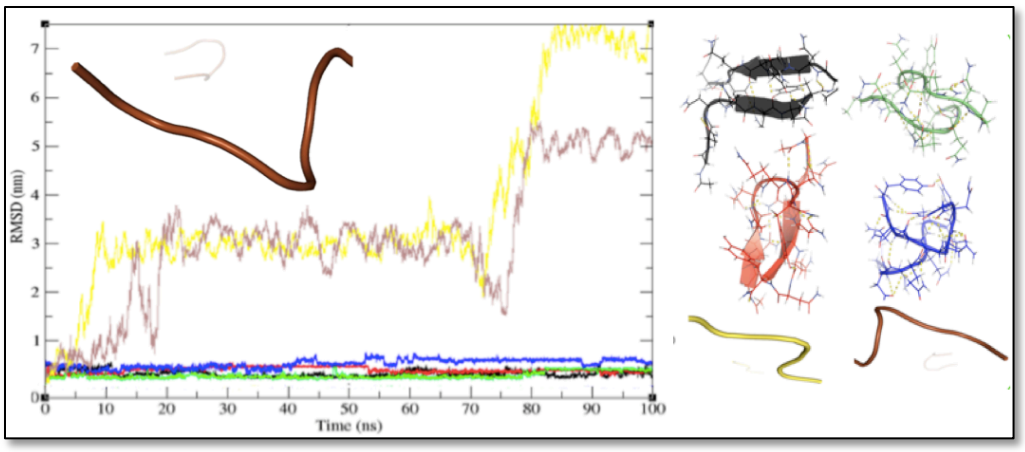


Figure S1: root mean square deviation (RMSD) for the **S1·S1** dimer. The first 100 ns of the 500 ns run of the simulation presented in the main text (Figure 1) is shown in **brown**. The kinetic stability of an **S1·S1** arrangement is tested, where side chain intermeshing provides the main force holding two peptides together. In two simulations (**yellow** and **brown**), the dimer completely unfolds and dissociates. In two other simulations (**green** and **blue**) the starting structure is largely maintained but the termini undergo significant changes to optimize the formation of a set of stabilizing contacts, which are different from those present in the starting structure. In two other structures (**red** and **black**) the starting **S1·S1** conformation transforms structurally into a pair of H-bonded, anti-parallel β-strands. These results suggest that the stabilizing contribution afforded by the intermeshing side chains is small. The RMSD values plotted are for both peptides in each dimer (and not to individual peptides).

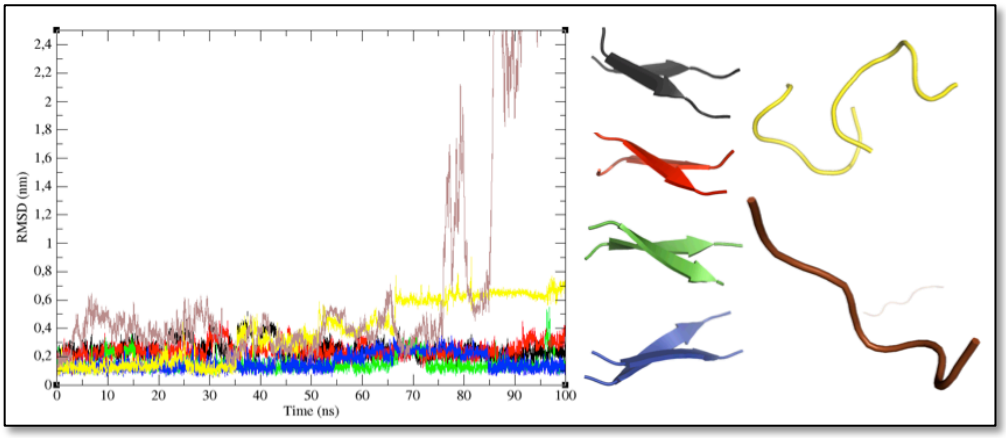


Figure S2: Root mean square deviation (RMSD) for the **S2** system. The **brown** trace represents the first 100 ns of the 500 ns MD run presented in the main text (Figure 1). The other trajectories correspond to repetitions of the simulations to test the convergence of the results. Interestingly, after 100 ns, four out of six **S2** dimers retain conformations closely resemble the starting structure. Two **S2** dimers significantly unfold after 60-80 ns, one of which, in **brown**, dissociates (note that the second **brown** monomer is in the distant background). As in Figure S1, the RMSD values plotted are for both peptides in each dimer, and not to individual peptides.

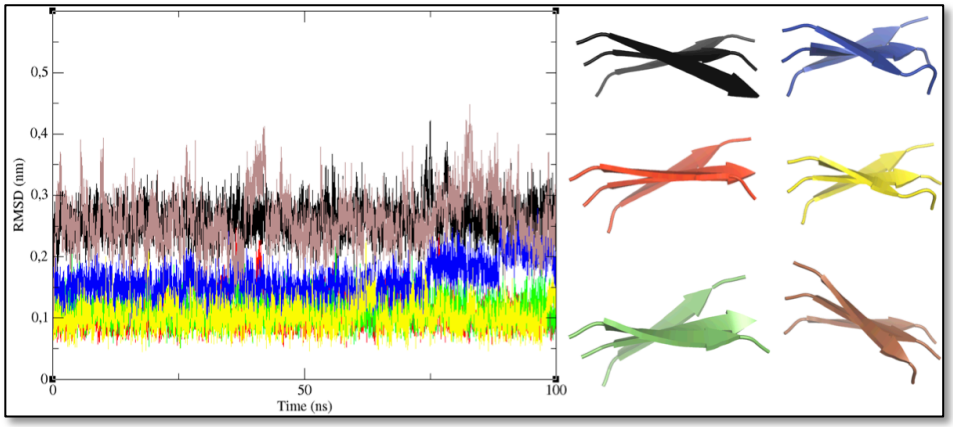


Figure S3: RMSD for the **S3 (which twists to become *S3* in the equilibrium stage)**. As in other supplementary figures, the **brown** run represents the first 100 ns of the 500 ns run shown in **Figure 1a** in the main text. In all the other independent MD calculations (run for 100 ns and shown in **black**, **blue**, **red**, **yellow** and **green**), the structures maintain a twisted three stranded β-sheet which minimal structural deviations. This is evidence that twisting provides extraordinary conformational stability. The reproducibility of the results corroborates our proposal that the ***S3*** structure is highly stable and acts to seed GNNQQNY fibril growth. The RMSD values plotted are for all three peptides in each trimer.

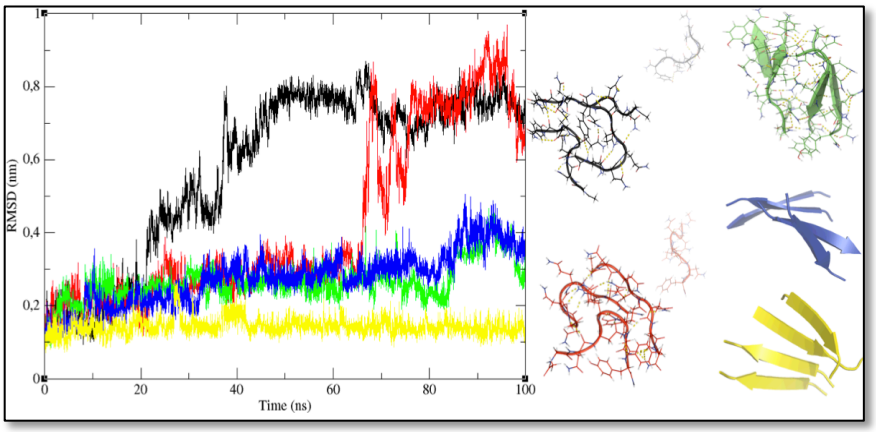
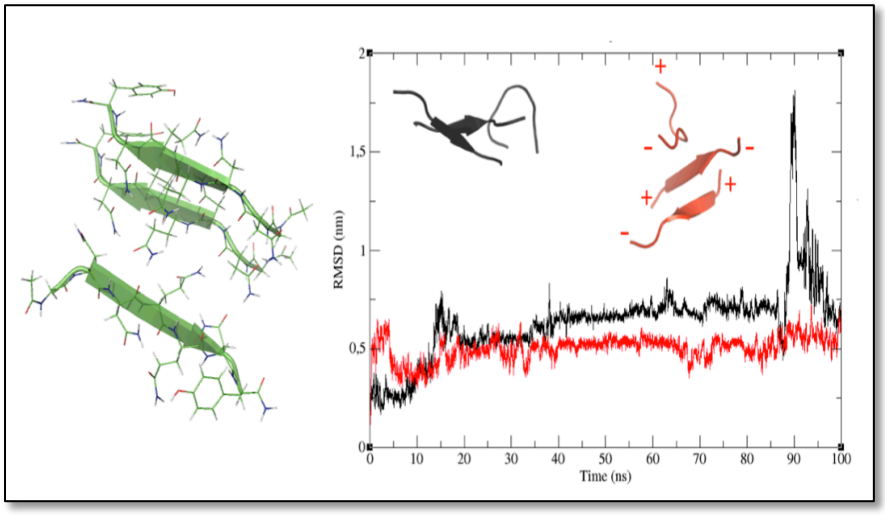


Figure S4: RMSD for the **S2·S2** system. Here the **black** trace corresponds to Figure 1b of the main text (*i.e.*, the first 100 ns of the 500 ns run). This structure is not stable enough to be considered as a key intermediate in Sup35 oligomerization, as can be seen below. All of the snapshots represent the last frame of the trajectory. Only in one case is the starting **S2·S2** configuration retained, although it is highly distorted (**green**). In two simulations (shown in **black** and **red**), one peptide dissociates and the structure evolves towards an ***S3***-like system, whereas in the other case (**blue**) it seems to convert into a four stranded β-sheet. The last frame of this four stranded β-sheet was subjected to an additional 100 ns of MD to test its kinetic stability (yellow). The sidechains of these two last systems are not shown to highlight their secondary structure, since they differ from the all-parallel β-sheet. These structures might dissociate into two ***S2*** systems and evolve along our proposed pathway into a seed for amyloid formation (see *Figure 3* in the main text). The RMSD values plotted are for correspond to the entire tetrameric assembly.



FigureS5: root mean square deviation (RMSD) for the **S2·S1** system. When the system is charged, its behavior is similar to that of **+S1·S1-** and when termini residues are neutral, it resembles **S2** trajectories. The RMSD values shown here, and the remaining figures, correspond to the whole system, not individual**.**

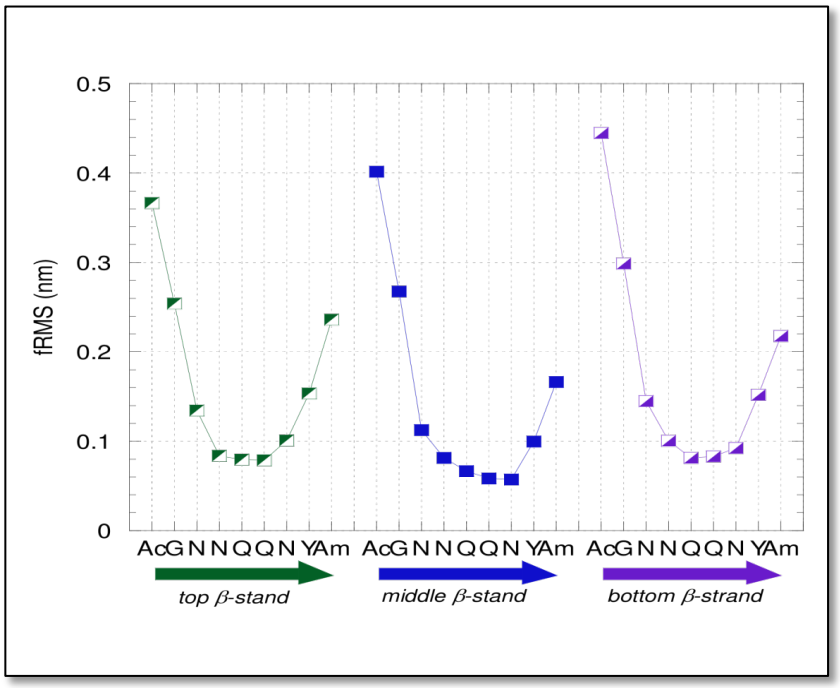


Figure S6: Root mean square fluctuation (RMSF) calculated for all the ***S3*** structures over the 500 ns trajectory. It is important to note that the lower values correspond to the residues that form the dry interface in the fibril. The fact that this low mobility originates when **S3** twists to yield ***S3*** suggests that side chain packing to yield the dry interface must proceed after ***S3*** forms, as can be seen in the proposed mechanism in *Supplementary Figure 11*.

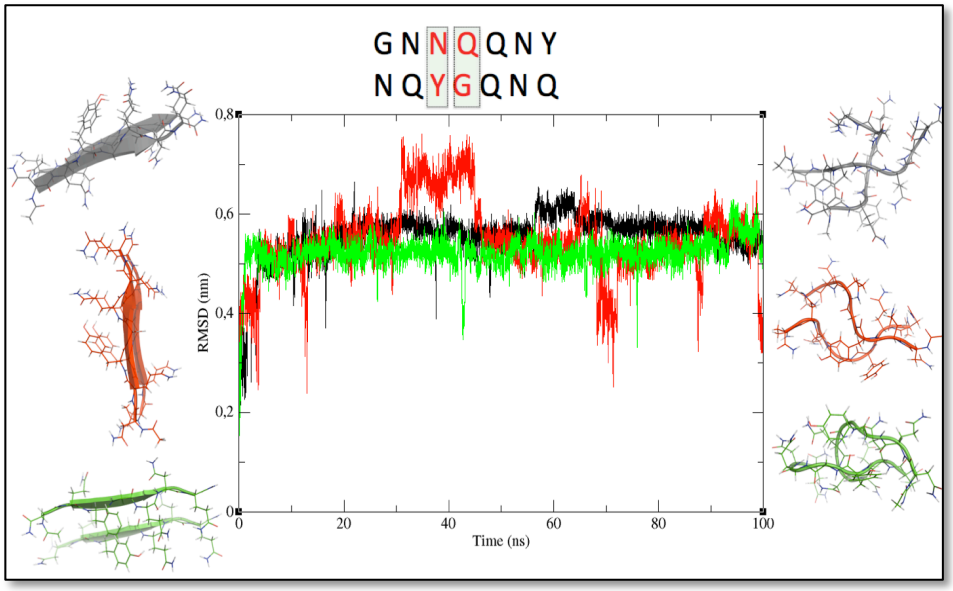


Figure S7: Root mean square deviation (RMSD) of the **NQYGQNQ** mutant. The wt sequence (GNNQQNY) was altered to break up the N-X-Q-X-N motif which had been proposed by Nelson *et al* (Nature, 2005)[1] to be key for dry interface stability. The substitutions N3Y and Q4G impede the formation of stable β-sheet structures, as gauged from three independent 100 ns simulations in which the RMSD value is higher than 0.5 nm. This is evidence that this force field is not biasing these Q/N rich sequences towards any particular conformation, and is therefore appropriate for this system.

Figure S8: Characterization of **GQNNGQNY** and **GNNQQNY** peptides by NMR. Removal of the central Q4 contacts results in non-aggregating systems.

The NMR chemical shift assignments were obtained by a standard approach utilizing TOCSY crosspeaks to assign intra-residue spin systems and ROESY crosspeaks to establish their sequential linkage. These ^{1}H assignments were corroborated and ^{13}C & ^{15}N assignments were obtained by analysis of the ^{1}H-^{13}C & ^{1}H-^{15}N HSQC spectra. Some medium range ROE crosspeaks were seen between the side chain Hβ and Hγ of Gln 5 and the Hδ and Hε of Tyr 7. In the case of the wild type sequence, we observed that three spin systems, belonging to Gln 4, Gln 5 and Asn 6, share a very similar chemical shift value for their ^{1}HN and are overlapped in the 1D ^{1}H spectra. The presence of three ^{1}HN groups here was confirmed by integrating the peaks in the 1D ^{1}H spectrum, and the ^{1}H-^{15}N HSQC spectrum. The ^{13}Cα, ^{13}Cβ, ^{15}N & ^{1}Hα chemical shift values (*vide infra*) measured for homologous ^{1}H in both peptides similar to each other. Once corrections for local sequence effects are applied, these values are very close to the standard values for expected for short unstructured peptides, and no significant tendency towards partial alpha or beta type backbone structure was detected. Some rather weak sequential $^{1}HN_{(i)}$ – $^{1}HN_{(i+1)}$ ROESY correlations were observed for both the peptides; these signals are consistent with small populations of helix or turn-like conformations. In addition, a few low intensity peaks arising from medium range contacts between the side chains of Gln 5 and Tyr 7 were also detected. These could arise when, for example, the backbone is in an extended conformation so that the Gln 5 and Tyr 7 side chains project in similar directions.

Based on these results, we conclude that these peptides, upon dissolving in aqueous solution, initially adopt a broad ensemble of conformations. This conclusion is consistent with the findings of a ^{1}H NMR study of GNNQQNY in 50% TFE.[38]

Sup35: GNNQQNY

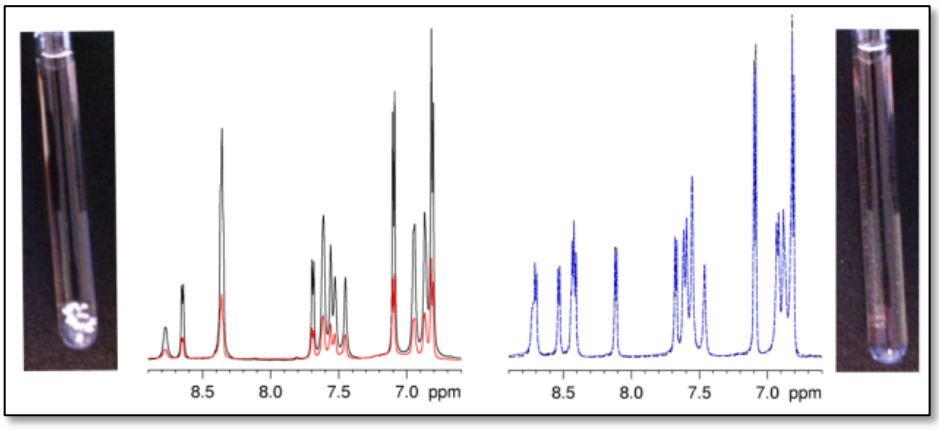
Residue	^{1}H^{15}N	^{1}H,^{13}Cα	^{1}H,^{13}Cβ	^{1}H ^{13}Cγ	^{1}Hδ12 or ^{1}Hε12 ^{15}Nδ or	^{1}Hδ22 or ^{1}Hε22 ^{15}Nε
Gly 1		3.86 **43.67**				
Asn 2	8.75 118.7	4.68 **53.50**	2.85, 278 **39.1 ± 0.2**		7.60† 112.8	6.94†
Asn 3	8.62 119.4	4.68 **53.84**	2.86, 2.78 **39.1 ± 0.2**		7.59† 113.0	6.93†
Gln 4	8.34 120.2	4.29 **56.32**	2.11, 1.98 **29.60**	2.33 **34.17**	7.50 112.5	6.84
Gln 5	8.333 120.1	4.28 **56.08**	1.97, 1.90 **29.79**	2.25 **34.04**	7.43 112.4	6.81
Asn 6	8.34 120.2	4.70 **53.49**	2.76, 2.65 **39.37**		7.53 112.8	6.86
Tyr 7	7.76 124.1	4.43 **59.00**	3.08, 2.90 **39.44**		7.10 (Hδ)	6.81 (Hε)

Sup35 Negative Control: GQNNGQNY

Residue	^{1}H^{15}N	^{1}H,^{13}Cα	^{1}H,^{13}Cβ	^{1}H ^{13}Cγ	^{1}Hδ12 - or ^{1}Hε12 ^{15}Nδ or	^{1}Hδ22 or ^{1}Hε22 ^{15}Nε
Gly 0		3.85 **43.6**				
Gln 1*	8.71 119.8	4.39 **56.41**	2.09, 1.99 **29.96**	2.35 **34.04**	7.54 112.3	6.86
Asn 2	8.69 120.1	4.71 **53.61**	2.85, 2.76 **39.13**		7.60 113.2	6.92
Asn 3	8.52 119.7	4.72 **53.65**	2.84 **39.05**		7.58 112.6	6.90
Gly 4*	8.42 108.9	3.93 **45.81**				
Gln 5	8.11 119.4	4.33 **55.82**	1.99, 1.90 **29.85**	2.25 **34.00**	7.45 112.4	6.82
Asn 6	8.40 120.3	4.72 **53.59**	2.79, 2.69 **39.34**		7.54 113.0	6.87
Tyr 7	7.66 125.0	4.40 **59.35**	3.08, 2.92 **39.6**		7.06 (Hδ)	6.78 (Hε)

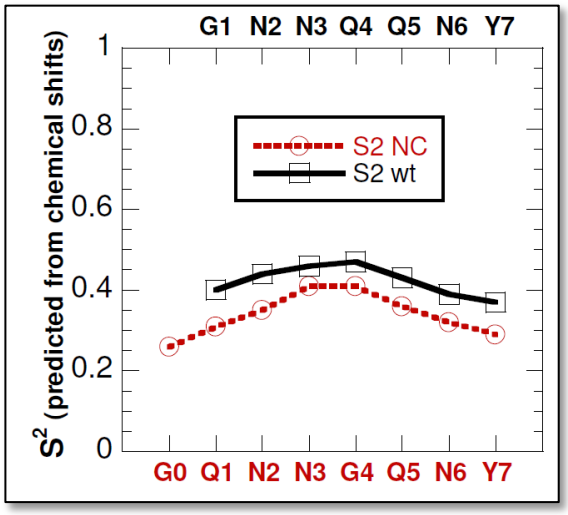
† Inter-residual ambiguity, *swapped relative to the wt sequence

The downfield region of the 1D ^{1}H NMR spectra of both peptides and these samples' photographs after five days of incubation are shown below:



Sup 35 GNNQQNY (*left panel*) 1 h (**black** spectrum) and 5 days (red spectrum & photo) after sample preparation. Sup 35 negative control (*right panel*) 1 h (dotted **black** spectrum) and 5 days (dashed blue spectrum and photo) after the sample was dissolved.

After five days of incubation at room temperature, visible aggregates appear in the GNNQQNY sample. Using the signal of DSS, as an internal intensity standard, we determined that 70% of GNNQQNY had aggregated after five days of incubation (red vs. black spectra, *vide supra*). In contrast, the control peptide showed no visible aggregates and no significant alternation in the NMR signal intensity after five days. These results are in line with the somewhat lower backbone order parameters for the negative control peptide with Gly 4 (*vide infra*) and support the conclusions drawn from the MD results on the importance of the central residues, particularly Gln 4 in wild type GNNQQNY oligomerization.



The predicted order parameter (S2) of wt (**black** open squares) and the neg. control peptide (red open circles).

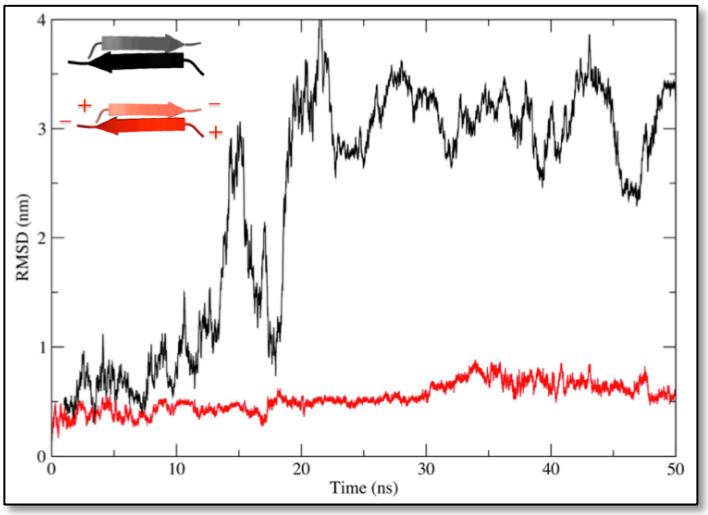


Figure S9: Comparing charged versus neutral termini in **S1·S1**. If the terminal residues are charged, the favorable electrostatic interactions increase the stability of this conformer. In contrast, the **S1·S1** system with capped, neutral termini quickly dissociates.

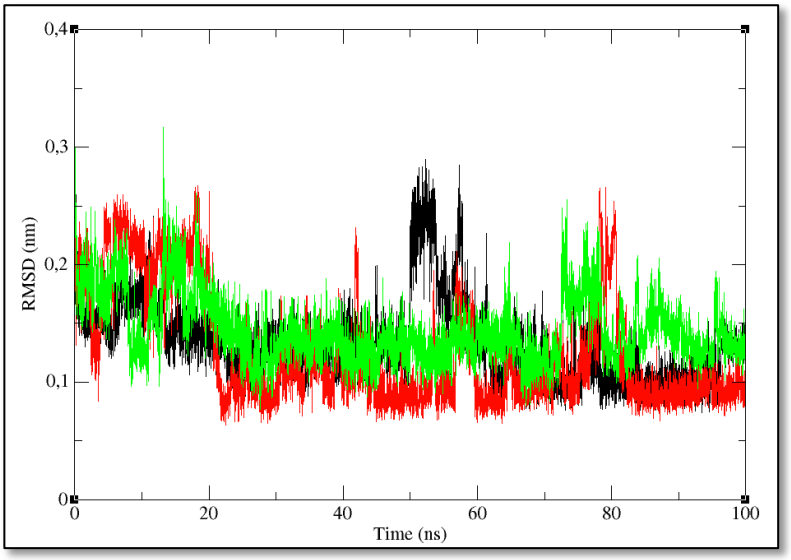


Figure S10: RMSD for the **S3·S2** system. Three different simulations were performed, varying the initial velocities on each atom (**red**) or the initial configuration (**green**), with respect to the initial one (**black**). In all cases, the RMSD was calculated with respect to the corresponding averaged structures.

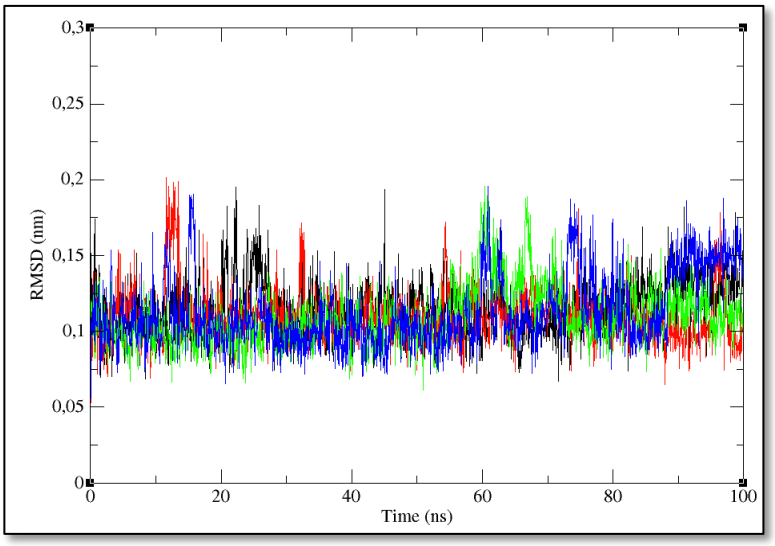


Figure S11: Stability of the trimers. The stability of the ***S3*** trimers that evolve towards combining to form the ***S3·S3*** hexamer (**Figure 4,** main text) were analyzed through their RMSD profiles. As expected from the multiple runs on ***S3*****-**like systems, they remained stable throughout the whole simulation period and that twisted conformation is retained when the dry interface is formed. The color code used is that corresponding to each of these trimers in Figure 4 in the main text.

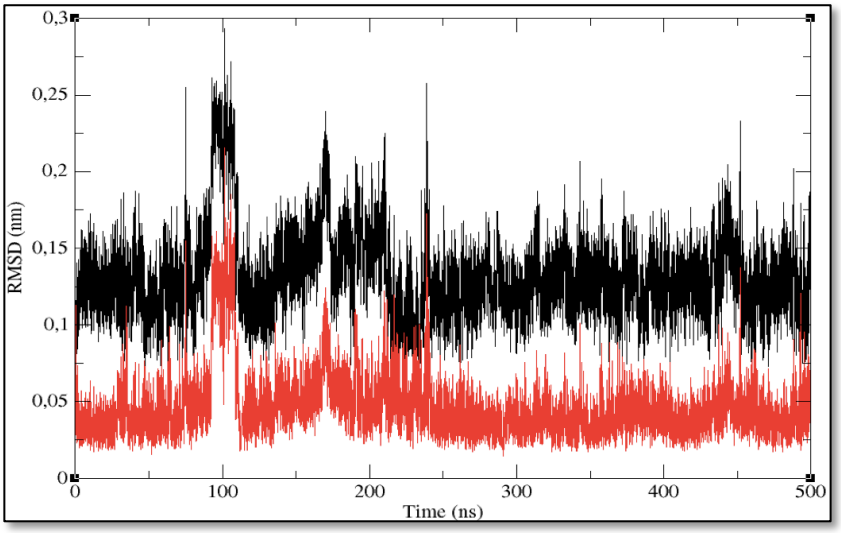


Figure S12: RMSD of the ***S3*** system fit to the averaged structure (**red**) and initial one (**black**), when the acetyl and amide terminal capping moieties are removed. *S3*'s remarkably stability is the reason for choosing this conformer for the ONIOM calculations: as there are no substantial variations during the whole trajectory, the averaged ensemble conformation is representative of all the frames.

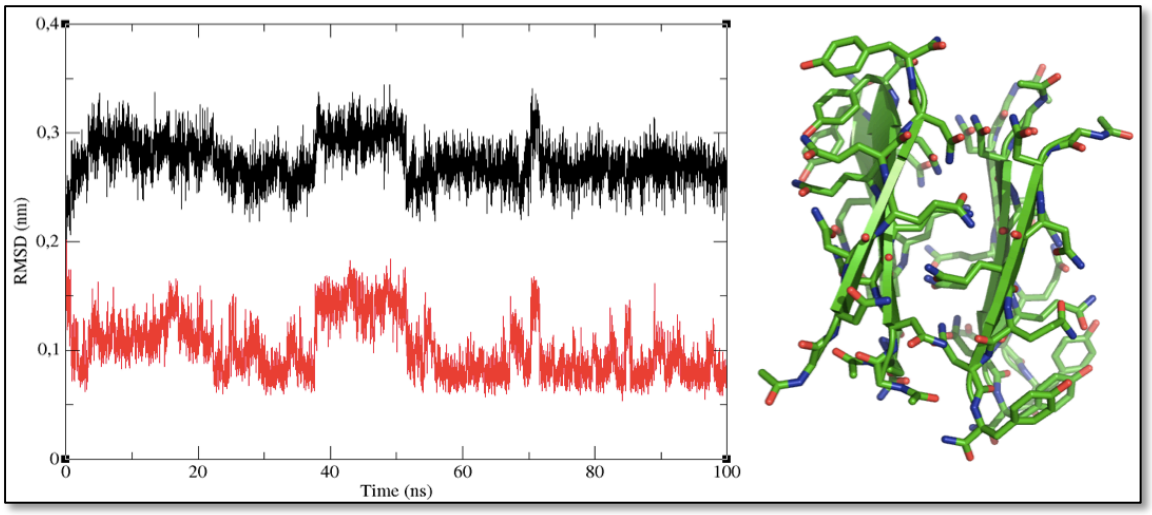


Figure S13: Stability of *S3·S3*. *Left*: The RMSD of the nucleus or seed for aggregation is shown below to illustrate how this system is properly represented by the force field and the set of parameters used in this study. The small fluctuations are mostly due to the acetyl and amide terminal moieties used to cap the termini. The structures of the resulting trajectory were fitted to two reference conformations: the initial one (at time zero, **black** curve) and the time-averaged conformation (**red** curve). *Right*: Cartoon and stick representation of the last frame of the 100 ns run, showing the twisted ***S3·S3*** system. Hydrogen atoms are not shown for the sake of clarity.

2.7.3 Supporting Tables

Table S1: Main findings in previous manuscripts analyzing GNNQQNY at atomic resolution.

Reference	Chief Methodology	Main Findings
(Balbirnie et al. 2001)[21]	Domain mapping / X-ray diffraction	Determined that a heptapeptide of the large Sup35 protein that is able to form long fibrils (as seen by EM) that bind amyloid-specific dyes. By X-ray diffraction, the spacings between the H-bonded (4.8 Å) and van der Waals packed (10 Å) peptides were determined.
(Nelson et al. 2005)[1]	X-ray crystallography	Determined the crystal structure of the Sup35 heptapeptide GNNQQNY with charged termini and a Zn^{++} coordinated to the C-terminal –COO- and N-terminal amine group. The structure revealed the formation of a complete set of side chain and backbone H-bonds. Each peptide forms a β-strand that is anti-parallel to the fibril axis. There is no twisting. The Sup. Materials of this paper provides ballpark estimates of the factors that contribute to the fibril's conformational stability, emphasizing the importance of H-bonding, close side chain packing, Tyr ring stacking and solvation, and suggest that structural nucleus could be formed by ≤ 6 peptides.
(Sawaya et al. 2007)[22]	X-ray crystallography	Determined the crystal structure of amyloid fibrils formed by several different short peptides. Whereas all the fibrils exhibit very tight sidechain packing and complete H-bond networks, there is a wide variation in the hydrophobicity, hydration and orientation of the strands.
(Tsemekhman et al. 2007)[16]	Energetics by Computation	First attempt to measure the stabilizing bonus that should arise from hyper-polarization of amide groups in the Sup35 amyloid fibril and the smaller oligomers that maintain the same conformation. Did not consider twisting.
(Guo and Eisenberg 2008)[23]	X-ray crystallography	Determined the structure of NNQQNY inserted in a folded protein. This segment adopts a twisted, antiparallel conformation that is incompatible with aggregation.
(Lipfert et al. 2005)[24]	MD	Observed that the release of water molecules drives oligomerization. They study the system with charged termini. They see more disorder at the termini when the peptides are aligned in parallel as compared to anti-parallel due to charge repulsion.
(Zheng et al. 2006)[25]	MD	Performed short (10 ns) simulations on small oligomers of different sizes based on the X-ray structure of Nelson et al. 2005. Peptides have terminal charges. Simulate point mutations. Recognize the importance of Q4, N6 & Y7 to stability. See that one sheet with 3 or more β-strands is stable. Propose that the nucleus could be as small as 3 or 4 peptides. Did not recognize the importance of twisting.
(Esposito et al. 2006)[26]	MD	Performed short simulations of longer oligomers of Sup35 in a fibril like conformation, starting from X-ray structure. Discovered that the fibril, composed of peptides with charged or neutral termini, undergoes left-handed twisting up to 10º. Report that the dry interface, esp. Q4, is very rigid. Twisting leads to further dry interface contacts.
(deSimone et al. 2008)[27]	MD	Utilized replica exchange MD to look at the stability of several small GNNQQNY oligomers. See that "S3·S3" and larger oligomers are stable, and that "S3·S2" is border-line. Emphasize that the exposed edge strands are "highly reactive".
(Vitagliano et al. 2008)[28]	MD	Utilized replicate exchange MD to determine the free energy surface for GNNQQNY dimers whose strands are in parallel or anti-parallel orientations, for charged, half-charged and neutral peptides. See that the energy differences between the two orientations are subtle. Neutralizing the terminal charges favors the parallel orientation.
(Zhang et al. 2007a)[29]	MD	Performed many simulations to follow the formation of GNNQQNY oligomers (dimers, trimers & tetramers). See twisting already in these small oligomers. The peptides seem to be charged. See that parallel β-sheets form faster and dissociate more slowly.
(Periole et al. 2009)[30]	MD	Determined that twisting is favored when the termini are neutral, and even more so when they are charged. Likewise, solvent water is not essential for twisting, but does stabilize the twisted form.

Table S1 (cont.): Main findings in previous manuscripts analyzing GNNQQNY at atomic resolution.

Reference	Chief Methodology	Main Findings
(Park et al. 2009)[31]	MD	Measured the energetics for adding a strand to a fibril, for several different peptides, including GNNQQNY. Find that the conformation adopted is determined by the energetics.
(van der Wel et al. 2007)[32]	ssNMR	Detected, via ssNMR measurements, the existence of polymorphism in Sup35 amyloid fibrils: monoclinic (like Eisenberg *et al.*'s structure) and orthorhombic. Emphasize the importance of Tyr ring packing.
(van der Wel et al. 2010)[33]	ssNMR	Reported three different types of fibrils; all are parallel, in-register β-sheets.
(Wang et al. 2008)[34]	MD	Utilized RT (298K) and higher temperatures (up to 498 K) to follow the dissociation of GNNQQNY duodecamers. Based on this, they propose a pathway for oligomer formation which emphasizes the importance of "S3·S1" and "S2·S2" tetramers as key intermediates/transition states. Peptides seem to have terminal charges and oligomers are twisted. Assigned equal importance to H-bonds and van der Waals interactions.
(Reddy et al. 2009)[35]	MD	Examined the process of the addition of GNNQQNY monomers initially in a random coil state onto a preformed amyloid fibril. Emphasized the importance of dehydration of GNNQQNY, a polar peptide, esp. compared to the same growth process for Aβ, a much more hydrophobic peptide.
(Reddy et al. 2010)[36]	MD	Computational studies of GNNQQNY beta sheets docking to form the dry interface. When two β-sheets, each formed of 8 monomers, dock together, water molecules can be trapped inside. These "water nanowires" are meta-stable for GNNQQNY but not seen for Aβ, which is much more hydrophobic.
(Marshall et al. 2010)[37]	Linear Dichroism, X-ray diffraction, Tyr fluorescence, acrylamide quenching	Monitored structural changes that occur when the Sup35 amyloid fibril crystallizes. Saw that the Tyr side chain undergoes important changes in conformation and exposure to solvent. Affirmed that the crystalline state is more stable than the fibril state.
(Portillo et al. 2012)[2]	ThT Fluor. AFM force spectrometry	Measured the effect of variation of pH and ionic strengthen on the kinetics of amyloid formation (monitored by ThT fluorescence) and the interaction energy of Sup35 peptides in a dimer (using AFM). Find that increasing the ionic strength and pH values close to the isoelectric point (5.3) speed amyloid formation, but that these environoment conditions do not strongly affect the interaction energy.
(Chae et al. 2004)[38]	Liquid state NMR	^{1}H NMR study of GNNQQNY in different concentrations of water&TFE. The obtained ^{1}Hα chemical shifts in 50% are characteristic of random coil.
(Nasica-Labouze *et al*. 2011)[39]	Coarse-grain implicit solvent MD	Coarse-grain, implicit solvent MD study of the different oligomers formed by GNNQQNY, starting from boxes with 3, 12 or 20 monomers. A large variety of oligomers and structures are observed. The most Representative structures are studied further using all atom MD with explicit solvent.
Nasica-Labouze & Mouseau 2012)[40]	Coarse-grain, implicit solvent MD	Coarse-grain, implicit solvent MD study of the kinetics of GNNQQNY association, starting from 20 monomers. Observe formation of a variety of beta sheet structures; with small oligomers forming before larger oligomers.

SYSTEM	ΔE_{int} (kCal/mol)
S3·S3	**-30.96**
S2·S3 -90ns	-5.68
S2·S3 -92ns	NOT CONSISTENT
S2·S3 -94ns	-7.74
S2·S3 -96ns	-11.85
S2·S3 -98ns	-4.13
S2·S3 -100ns	-12.70
$<\Delta E_{int}$ (S3·S2) >	**-8.42**

FUNCTIONAL	GAIN OF ΔE_{int} UPON TWISTING (kCal/mol)	
B3LYP	-8.92	(-12.88)
M06-2X	-9.11	(-12.98)

Table S2: Summary of the ONIOM (DFT:PM6) results.

All the values were computed with the PCM method to account for environment effects. (*Left*) Values obtained per structure considered. In the case of S3·S3, the low RMSD obtained during the long simulation times allows for a direct estimation of ΔEint from the time-averaged structure. For the S3·S2 system, snapshots from the last 10 ns of the trajectory were extracted. The ΔEint value was calculated for each individual structure and averaged.

(*Right*). The advantage of twisting is evidenced by a gain in interaction energy. To check these results' consistency we compared values obtained with a continuum solvation model with those from gas-phase calculations -in parenthesis- and found an excellent agreement between both the B3LYP and M06-2X functionals.

Table S3: Summary of the simulation results reported here.

Simulation	Simulations X Length (ns)	Main results	Corresponding Figure
S1·S1	1 x 500 5 x 100 1 x 50 (charged)	Stable for no longer than 10 ns. Dissociation or rearrangement into anti ‖ β-sheet	1 (main text), 1 (Sup. Info) & 9 (Sup. Info)
S2	1 x 500 5 x 100	Stable for at least 75 ns. Remarkably stability	1 (main text) & 2 (Sup. Info)
S3	1 x 500 5 x 100	Remarkably stable in all of the cases	1 (main text) & 3 (Sup. Info)
S2·S2	1 x 500 5 x 100	Evolution towards S3 / Structural rearrangement	1 (main text) & 4 (Sup. Info)
S2·S1	1 x 100 (charged) 1 x 100 (capped)	When charged, behaves as ***+S1·S1-***. When capped, behaves as ***S2***.	5 (Sup. Info)
NQYGQNQ	3 x 100	β-sheet secondary structure not retained	7 (Sup. Info)
S3·S2	3 x 100	Stable, although not as remarkable as S3·S3	10 (Sup. Info)
4 x *S3*	1 x 100	Evolution towards *S3·S3* nucleus	4 (main text) & 11 (Sup. Info)
S3·S3	1 x 100	Stable	12 (Sup. Info)

Total simulation time: 5.05 μs.

Table S4: Semi-quantitative analysis of the number of H-bonds required to overcome the electrostatic repulsion between charged peptides arranged in a two-stranded parallel β-sheet.

System	Number of H-bonds (Backbone + side chain = total)	Stable?
G-N-Y **G-N-Y**	2 + 1 = 3	NO
G-N-N-Y **G-N-N-Y**	3 + 2 = 5	NO
G-N-N-Q-Y **G-N-N-Q-Y**	4 + 3 = 7	YES
G-N-N-Q-Q-Y **G-N-N-Q-Q-Y**	5 + 4 = 9	YES
G-N-N-Q-Q-N-Y **G-N-N-Q-Q-N-Y**	6 + 5 = 11	YES

2.7.4 Supporting References

(1) Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D (2005) Structure of the cross-beta spine of amyloid-like fibrils *Nature* 435, 773-778.

(2) Portillo AM, Krasnoslobodtsev AV, Lyubchenko YL (2012) Effect of electrostatics on aggregation of prion protein Sup35 peptide *J. Phys.: Condens. Matter*. 24: 164205.

(3) Hess B Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput*. 4: 435.

(4) Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Kepleis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78: 1950-8.

(5) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. 79: 926-935.

(6) Berhanu WM, Hansmann UHE (2012) Side-chain hydrophobicity and the stability of Abeta 16-22 aggregates. *Protein Science* 21: 1837-1848

(7) Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N·lon(N) method for Ewald sums in large systems. *J. Chem. Phys*. 98: 10089-10092

(8) Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys*. 81: 3684.

(9) Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem*. 18: 1463-1472.

(10) Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys*. 52: 255

(11) Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys*. 52: 7182.

(12) Lange OF, van der Spoel D, de Groot BL (2010) Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J*. 99: 647-55.

(13) Dapprich S, Komaromi I, Byun KS, Morokuma K, Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part 1. The calculation of energetics, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: TEOCHEM* 461-462: 1.

(14) Stewart JJ (2007) Optimization of parameters for semi empirical methods V: modification of approximations and application to 70 elements. *J Mol. Model* 13: 1173-1213.

(15) Gesto DS, Cerqueira NM, Fernandes PA, Ramos MJ (2013) Unraveling the enigmatic mechanism of L-asparaginase II with QM/QM calculations. *J. Am. Chem. Soc*. 135: 7146-58.

(16) Tsemekhman K, Goldschmidt L, Eisenberg D, Baker D (2007) Cooperative hydrogen bonding in amyloid formation. *Protein Sci*. 116: 761-4.

(17) Cerqueira N, Fernandes P, Ramos MJ (2011) Computational mechanistic studies addressed to the transimination reaction present in all pyridoxal 5' phosphate-requiring enzymes. *J. Chem. Theory Comput*. 7: 1356-1368.

(18) Martín-Pintado N,Yahyaee-Anzahaee M, Deleavey GF, Portella G, Orozco M, Damha MJ, González C (2013) Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc*. 135: 5344-7.

(19) Zhao Y, Truhlar D (2008) The M06 suite of density functional for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functional and 12 other functionals. *Theor. Chem. Acc*. 120: 215-241.

(20) Frisch MJ *et al*. Gaussian 09 Revision A.2, 2009. Gaussian Inc. Wallingford CT.

(21) Balbirnie M, Grothe R, Eisenberg D (2001) An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. (USA)* 98: 2375-2380.
(22) Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen AO, Riekel C, Eisenberg D (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447: 453-457.
(23) Guo Z, Eisenberg D (2008) The structure of a fibril-forming sequence NNQQNY, in the context of a globular fold. *Prot. Sci.* 17: 1617-1623.
(24) Lipfert J, Franklin, J, Wu F, Doniach S. (2005) Protein misfolding and amyloid formation for the peptide GNNQQNY from yeast prion protein Sup35: simulation by reaction path annealing. *J. Mol. Biol.* 349: 648-651.
(25) Zheng J, Ma B, Tsai CJ, Nussinov R (2006) Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *Biophys J* 91: 824-833.
(26) Esposito L, Pedone C, Vitagliano L (2006) Molecular dynamics analyses of cross-beta-spine steric zipper models: beta sheet twisting and aggregation. *Proc. Natl. Acad. Sci. (USA)* 103: 11533-11538.
(27) deSimone A, Esposito L, Pedone C, Vitagliano L. (2008) Insights into the stability and toxicity of amyloid-like oligomers by replica exchange molecular dynamics analyses. *Biophys J.* 95: 1965-1993.
(28) Vitagliano L, Esposito L, Pedone C, DeSimone A. (2008) Stability of single sheet GNNQQNY aggregates analyzed by replica exchange molecular dynamics: antiparallel versus parallel association. *Biochem Biophys Res Com* 377: 1036-1041.
(29) Zhang Z, Chen H, Bai H, Lai L. (2007) Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35. *Biophys. J.* 93: 1484-1492.
(30) Periole X, Rampioni A, Vendruscolo M, Mark AE (2009) Factors that affect the degree of twist in beta-sheet structures: A molecular dynamics simulation study of a cross-beta filament of the GNNQQNY peptide. *J. Phys. Chem.* 113: 1728-1737.
(31) Park J, Kahng B, Hwang W (2009) Thermodynamic selection of steric zipper patterns in the amyloid cross-beta spine. *PLoS Comput. Biol.* 5: e1000492.
(32) van der Wel PCA, Lewandowski JR, Griffin RG (2007) Solid state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J. Am. Chem. Soc.* 129: 5117-5130.
(33) van der Wel PCA, Lewandowski JR, Griffin RG (2010) Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR *Biochemistry* 49: 9457-9469.
(34) Wang J, Tan C, Chen HF, Luo R (2008) All-atom computer simulations of amyloid fibrils disaggregation. *Biophys. J.* 95: 5037-5047.
(35) Reddy G, Straub JE, Thirumalai D (2009) Dynamics of locking of peptides onto growing amyloid fibrils. *Proc. Natl. Acad. Sci. (USA)* 106: 11948-11953.
(36) Reddy G, Straub JE, Thirumalai D (2010) Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl. Acad. Sci. (USA)* 107: 21459-21464.
(37) Marshall KE, Hicks MR, Williams TL, Hoffmann SV, Rodger A, Dafforn TR, Serpell LC (2010) Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: structural changes accompany a fiber-to-crystal switch. *Biophys. J.* 98: 330-338.
(38) Chae YK, Lee K, Kim Y (2004) NMR Studies of the prionogenic peptide derived from Sup35 Protein. *Prot. Pept. Lett.* 11: 23-28.
(39) Nasica-Labouze J, Meli M, Derreumaux P, Colombo G, Mousseau N (2011) A multiscale approach to characterize the early steps of the amyloid-forming peptide GNNQQNY from the yeast prion Sup-35. *PLoS Comp. Biol.* 7: e1002051.

Chapter 3

Structural Characterization of the Minimal Segment of TDP-43 Competent for Aggregation

We build too many walls and not enough bridges.
Isaac Newton

3.1 Abstract

TDP-43 is a nuclear protein whose abnormal aggregates are implicated in ALS and FTLD. Recently, an Asn/Gln rich C-terminal segment of TDP-43 has been shown to produce aggregation *in vitro* and reproduce most of the protein's pathological hallmarks in cells, but little is known about this segment's structure. Here, CD and 2D heteronuclear NMR spectroscopies provide evidence that peptides corresponding to the wild type and mutated sequences of this segment adopt chiefly disordered conformations that, in the case of the wild type sequence, spontaneously forms a β-sheet rich oligomer. Moreover, MD simulation provides evidence for a structure consisting of two β-strands and a well-defined, yet non-canonical structural element. Furthermore, MD simulations of four pathological mutations (Q343R, N345K, G348V and N352S) occurring in this segment predict that all of them could affect this region's structure. In particular, the Q343R variant tends to stabilize disordered conformers, N345K permits the formation of longer, more stable β-strands, and G348V tends to shorten and destabilize them. Finally, N352S acts to alter the β-stand register and when S352 is phosphorylated, it induces partial unfolding. Our results provide a better understanding of TDP-43 aggregation process and will be useful to design effectors capable to modulate its progression.

3.2 Introduction

TDP-43 is a 414 residue protein involved in mRNA processing and transport to the cytoplasm[1,2]. It has two RRM RNA binding domains and the N-terminus carries motifs for nuclear localization as well as nuclear export. These motifs allow TDP-43 to return to the nucleus after transport of mRNAs. The C-terminal half of TDP-43 contains an Asn/Gln rich region as well as a Gly rich region and has been suggested to contain a prion-like domain[3].

In 2006, TDP-43 was discovered to be the main protein component of abnormal protein aggregates in the cytoplasm of nerve cells in FTLD and ALS[4,5]. In these aggregates, TDP-43 molecules are ubiquinated, hyperphosphorylated and often truncated to yield 23-27 kDa C-terminal fragments, with the nuclei of cells harboring abnormal TDP-43 cytoplasmic aggregates becoming deficient in TDP-43[6].

Many genetic, cell biological and biochemical studies have provided evidence implicating TDP-43 in neurological disorders[2]. In particular, TDP-43 mutations associated with early disease onset or poor prognosis are linked to increased aggregation of the protein[7]. However, how TDP-43 harms cells is still an open question. The lack of TDP-43 in the nucleus could be detrimental to the proper splicing and transport of mRNAs. Alternatively, the hyperphosphorylated, ubiquitinated aggregates could be cytotoxic. So far, results have been reported supporting both these mechanisms, potentially making them both pathologically relevant[2].

Among the various functional domains present in the TDP-43 protein, one of the best characterized regions is represented by the Asn/Gln-rich segment spanning residues 321-366 in the C-terminus of this protein. In particular, this region has been previously described to be involved in the interaction with hnRNP proteins[8], polyglutamine repeats[9], and nuclear "Gems"[10]. In addition to all these potential connections, tandem repetitions of this Asn/Gln rich region of TDP-43 are capable of aggregating in cells and of recruiting normal TDP-43 into these aggregates[11]. Furthermore, the aggregated TDP-43 displays modifications that resemble most of the pathological hallmarks associated with the full length protein such as phosphorylation and ubiquitination. Within this region, there are three short segments that are rich in Asn and Gln residues, especially the first two (**Fig.1**). When the residues of any of these short segments are substituted by alanines, the resulting variant is incapable of forming aggregates[11]. On the basis of these results, this region, referred to as TDP-43 (342-366) was identified as a minimal TDP-43 segment capable of recapitulating its aggregation process[11].

Approximately forty TDP-43 mutations have been implicated in hereditary cases of ALS or FTLD; an up to date list can be found at https://www.molgen.ua.ac.be/FTDMutations. Almost all these mutations map to the sixth TDP-43 exon, which includes the minimal TDP-43 aggregative segment. In particular, the pathogenic mutations Q343R[12], N345K[13], G348V[14] and N352S[15,16] have been identified in the 321-366 region, are associated with early-onset familial ALS without dementia, and usually show dominant-type inheritance.

At present, little information is available regarding the consequences of these mutations on TDP-43 biological properties. Recent studies have shown that the G348V and N352S mutations do not affect RNA splicing but do increase aggregation within cells[17]. The conservation of the wild type amino acid residue from man to fish for Q343, N345 and G348 suggests that these residues are crucial; in contrast, N352 is replaced by Ser in fish[14], which suggests that this position may be more tolerant to substitution. Secondly, it has also been recently established that pathological missense substitutions positions Q343R, G348C, and N352S considerably increase the half-life of the TDP-43 protein[18]. These findings prompted us to explore whether the TDP-43 341-366 region could form a structural nucleus for aggregation. To meet this objective, we have characterized peptides corresponding to the WT TDP-43 sequence and point mutants Q343R, N345K and G348V using NMR and CD spectroscopies. The results show that these peptides adopt an unfolded ensemble and that wild type sequence can spontaneously self-associate to form a β-sheet rich oligomer that strongly enhances ThT fluorescence.

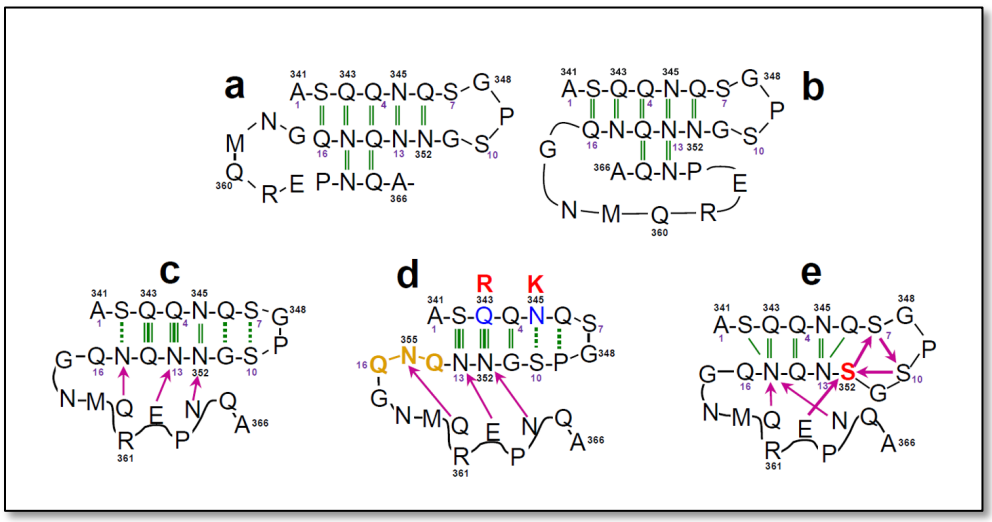


Figure 1. Schematic Structure of the TDP-43 (341-366) polypeptide. Schematic diagrams showing the configuration of the 341-366 Asn/Gln rich region of TDP-43: **(a)** arranged in an anti-parallel β-sheet. **(b)** Configured as a mixed anti-parallel/parallel β-sheet. **(c)** The structure observed during the course of MD for the wild type sequence. **(d)** The Q343R and N345K variants, where the out-of-register Q354-N355-Q356 segment is colored **gold**. **(e)** The N352S variant, where Ser hydroxyl interactions are represented as **purple** arrows. The positions of the wild type residues of the point mutations studied here are colored **blue** and the mutated residues are colored **red**. Some residues are numbered according to their position in the full length TDP-43 protein (**black**) or this segment (**bluish-purple**), *i.e.* A341 = 1, S342 = 2, Q343 = 3, *etc*. Interactions linking β-strand 2 and the C-terminal segment are highlighted with **purple** arrows. The position of β-secondary structure observed during the MD simulation (panels **c**, **d** & **e**) is marked in **green**: three lines ≡ almost always present; two green lines = usually present; dotted single line — sometimes present.

We hypothesize that this segment of TDP-43 as a monomer could occasionally adopt a β-sheet rich structure that may serve as a nucleus for oligomerization. To characterize this conformation, whose population is too low for our experimental methods, we have modeled TDP-43 (341-366) as a three stranded β-sheet and studied its structure and stability by molecular dynamics (MD) methods. In parallel, we have also studied how different pathologically relevant mutations; namely, Q343R, N345K, G348V and N352S can affect the structure and stability of this β-sheet. The results obtained have been discussed bearing in mind that mature TDP-43 aggregates are structurally diverse and that the aggregation process will involve interactions with other regions of the full length TDP-43 protein, as recently reported[19,20].

3.3 Results

Experimental Results

The backbone regions of the 2D ^{1}H-^{15}N, TOCSY & NOESY and ^{1}H-^{13}C HSQC spectra of the Q343R variant of TDP-43 (residues 342-366) are shown in **Fig. 2**. Despite the low chemical shift dispersion and the presence of several Asn and Gln residues, all the resonances could be uniquely assigned except for the ^{15}N of Q354 and Q356. Similar spectra were obtained and assigned for the WT, N345K and G348V variants (**Sup. Fig. 1a**). Their chemical shifts are chiefly similar to those of Q343R, except for the substituted residues and for the residues bordering the substitution. The complete assignments are listed in Table 1 in the **Supplementary Information**. The sidechain resonances were also completely assigned (**Sup. Fig. 2**). The chemical shift difference of the $^{13}C\beta$ and $^{13}C\gamma$ is consistent with the *trans* conformation for the P349 and P363 peptide bond. The chemical shift of M359's Hγ is consistent with the sulfur being reduced. One interesting difference is that the Hδ of Pro 349, which are degenerate in WT, Q343R and N345K, have unique chemical shift values in G348V (**Sup. Fig. 2**). Therefore, these H adopt unique magnetic environments when G348 is replaced by V, which could reflect a stiffening of the backbone.

The $^{1}H\alpha$, $^{13}C\alpha$, $^{13}C\beta$ and ^{1}HN chemical shifts also provided residue level information on secondary structure content[21] and dynamics[22]. The analysis of these chemical shifts, using TALOS+[23], indicated that these peptides are moderately flexible disordered ensembles (**Fig. 2d**). For example, most $^{13}C\alpha$ deviated <0.2 ppm from the expected coil values[21] (**Fig. 2d**). In fully populated α-helices or β-strands, deviations of +3.1 and -1.5 ppm, respectively, are observed[24]. The scarcity of stable H-bonded structures in our samples was also corroborated by ^{1}HN temperature factors of < -5 ppB·K^{-1} (**Sup. Fig. 3**)[25]. Nevertheless, we cannot rule out small populations (≤ 10%) of ordered conformations, such as the β-hairpin characterized by MD (see below). The flexibility of the polypeptide chain was predicted to be highest at the termini and lowest around residues 347-349 and 361-363 (**Fig 2d**). Finally, the rigidity of the 347-349 segment is somewhat more apparent in the G348V variant.

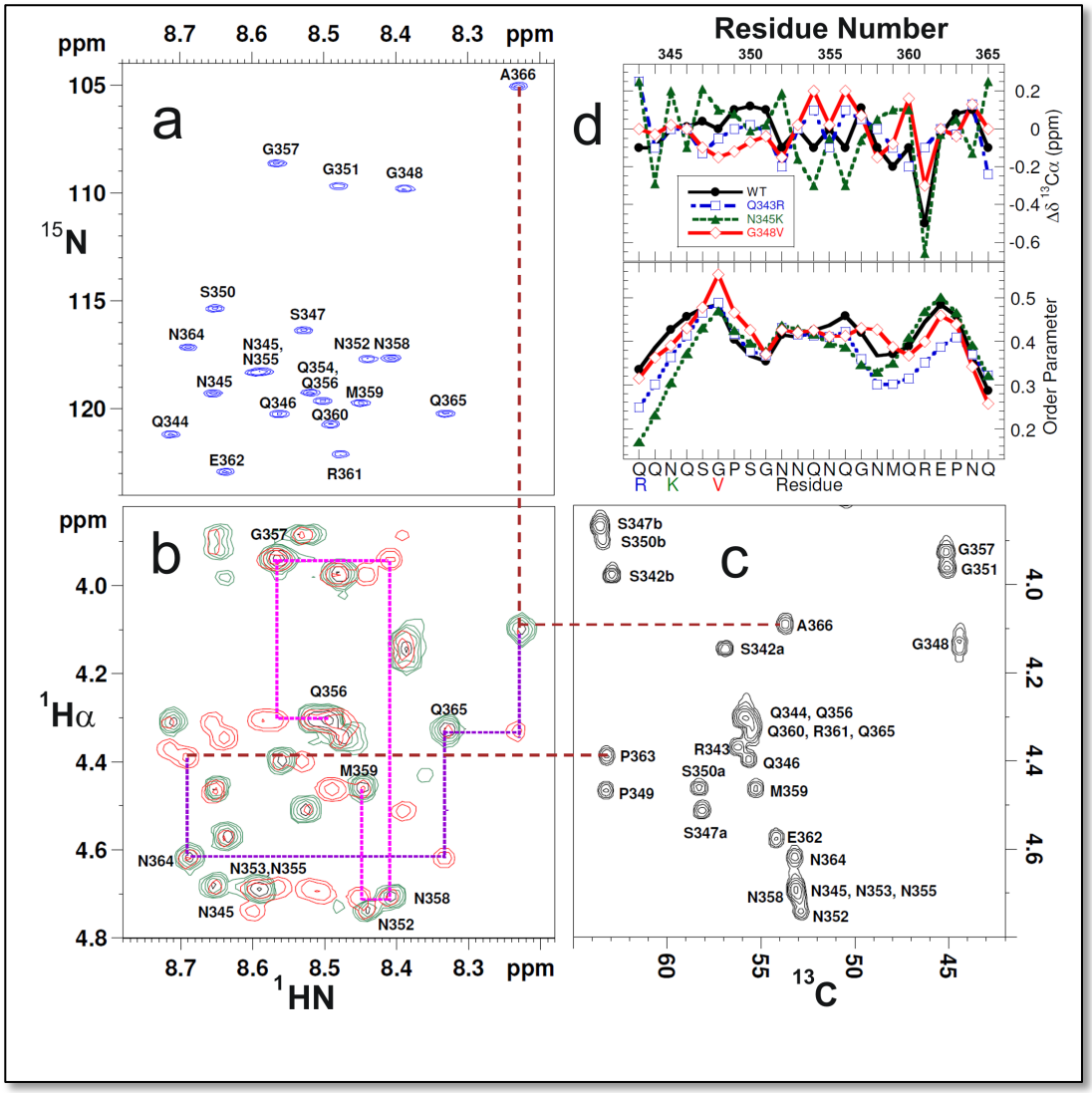


Figure 2: NMR Structural Characterization of TDP-43 (342-366) Peptides. **(a)** Backbone region of the ^{1}H-^{15}N HSQC spectrum (blue peaks). All ^{1}H-^{15}N signals are labeled. Note that the peak of A366 is folded; its correct ^{15}N chemical shift is 131.0 ppm. **(b)** The HN-Hα regions of the 2D ^{1}H TOCSY (green, 60ms mixing time) and NOESY (red, 150 ms mixing time) spectra. **(c)** The ^{1}H-^{13}C HSQC spectrum (**black** peaks). These spectra were recorded at pH 6.38 & 5 ºC. Brown dashed lines connect some signals which link nuclei from the same residue. Within the TOCSY/NOESY panel, two representative series of sequential inter-residual NOESY connectivities are shown as **purple** or **magenta** dotted lines. **(d)** ***Top panel*:** The conformational ^{13}Cα chemical shifts ($\Delta\delta^{13}$Cα) for TDP-43 (342-366) peptides. Values of +3.1 ppm and -1.5 ppm are expected for fully populated α-helices and β-sheets, respectively. The values observed here are much smaller and point to a low population of ordered conformations. The N- and C-terminal residues are not shown. ***Bottom panel*:** The predicted order parameter (S^2) for TDP-43 (342-366) peptides. Values of one indicate rigidity; values of zero indicate fully flexible chains. The estimated uncertainty in these values is 0.05.

In NOESY spectra, $H\alpha_{i-1}$ – HN_i resonances could be clearly assigned for all observed, non-overlapped signals. Strong $H\alpha_{i-1}$ – Hδ Pro NOEs were observed which corroborate that the Pro are mainly in the *trans* conformation. The absence of medium and long range NOEs is consistent with a lack of stable hydrophobic contacts and β-structure. There is one HN_{i-1} – HN_i NOE signal between N364 – Q365 seen in all spectra, and S347-V348, G357-N358 and R361-E362 crosspeaks are present in the G348V variant's spectrum (**Sup. Fig. 4**). These NOEs are indicative of turns.

Whereas the chemical shift values of the resonances of the WT peptide were similar to those of the other variants, it was evident that the WT peptide produces poorer quality NMR spectra (**Sup. Fig. 1**). In fact, the 1D ^1H spectra recorded before and after acquisition of the TOCSY and NOESY spectra (16 hour difference) showed a decrease of approximately 75% in the peptide signal intensity (**Sup. Fig. 5a).** The observation of a visible haze in this sample (**Sup. Fig. 5b**), suggested that the WT peptide had self-associated to form slowly tumbling aggregates that are invisible to liquid state NMR.

To corroborate the NMR results obtained on the variants and to characterize the NMR-invisible WT oligomers, we turned to far-UV CD spectroscopy. At pH 6.8, 25ºC, the variants Q343R, N345K and G348V showed similar CD spectra with minima near 198 nm that are typical of disordered ensembles (**Fig. 3a**). In contrast, a freshly prepared wild type peptide yielded a somewhat different CD spectrum with increased negative signal near 220 nm; and an aged WT sample showed a remarkably different spectrum with a minimum at 218 nm and a maximum at 199 nm. Similar results were obtained from additional experiments preformed at pH 8.2, 25º; pH 6.8, 5ºC and pH 8.2, 5ºC. Quantitative algorithmic analysis of these spectra shows that the Q343R, N345K and G348V variants and freshly dissolved WT peptide are dominated by turn and coil conformations. There is a remarkable increase in β-structure and decrease in coil in the aged WT sample. Whereas the freshly dissolved WT peptide contains 23% β-structure and 56% coil, the aged WT sample possesses 44% β-structure and 28% coil (**Sup. Table 2**).

Based on these observations, we then prepared another sample of WT TDP-43 (342-366) and recorded a series of CD spectra over several hours. These spectra revealed that peptide transforms structurally from an unfolded ensemble to conformations rich in β-structure (**Fig. 3b**). The observation of an isodichroic point, near 210 nm, is consistent with there being two main conformational ensembles, one disordered and one rich in β-structure. This experiment was repeated once and similar results were observed (data not shown).

It is surprising that the WT peptide oligomerized in solution conditions whereas the pathological variants Q343R, N345K and G348V did not show this tendency. The WT peptide has a net charge of +0 and the Q343R and N345K carry a net charge of +1 at pH 6.8. Electrostatic repulsion can significantly affect protein solubility[26] and aggregation[27]. Therefore, we recorded additional CD spectra of the peptides in increasing concentrations of KCl to test if aggregation occurs when charge – charge repulsion is screened. Such a condition would be more reflecting the physiological conditions, where high macromolecule and metabolic concentrations produce a crowding effect that promotes the association of peptides and proteins[28]. We then used CD spectroscopy to test if TDP-43

peptide variants oligomerize under crowding conditions. However, neither high salt nor crowding conditions were found to promote self association and structure formation in the Q343R, N345K and G348V variants (**Sup. Fig. 6**).

Next, ThT fluorescence was employed to obtain insight into the nature of the oligomers formed by the wild type TDP-43 (342-366) peptide. The variants Q343R, N345K and G348V and also RNase A, a well-folded monomeric protein, did not enhance ThT fluorescence. In contrast, the enhancement of the ThT fluorescence provoked by the WT TDP-43 peptide -over 20 fold- was almost as strong as that caused by the amyloid Aβ polypeptide implicated in Alzheimer's disease (**Fig. 3c**). It is interesting that aggregating tandem repeats of the 342-366 segment do not enhance ThS fluorescence *in vivo* (unreported observations), which supports the idea that the context of this segment within the complete protein versus isolated in a peptide, or in cell versus *in vitro* is key to its aggregates' superstructure.

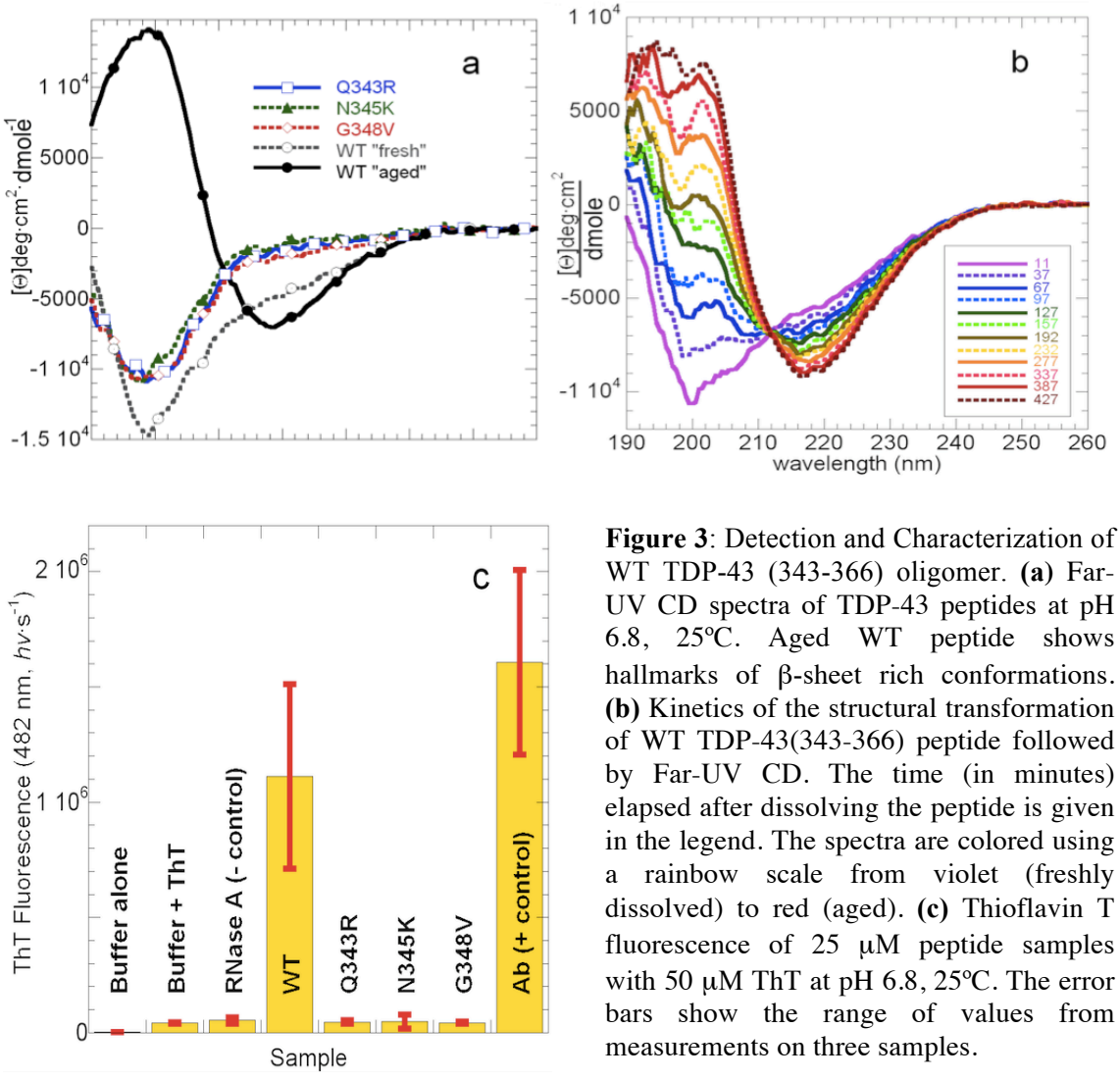


Figure 3: Detection and Characterization of WT TDP-43 (343-366) oligomer. **(a)** Far-UV CD spectra of TDP-43 peptides at pH 6.8, 25ºC. Aged WT peptide shows hallmarks of β-sheet rich conformations. **(b)** Kinetics of the structural transformation of WT TDP-43(343-366) peptide followed by Far-UV CD. The time (in minutes) elapsed after dissolving the peptide is given in the legend. The spectra are colored using a rainbow scale from violet (freshly dissolved) to red (aged). **(c)** Thioflavin T fluorescence of 25 μM peptide samples with 50 μM ThT at pH 6.8, 25ºC. The error bars show the range of values from measurements on three samples.

Finally, we considered the possibility that a peptide corresponding to just the segment 341 – 357, which is particularly rich in Asn and Gln (**Fig. 1**), might form a preferred conformation. Although this peptide is sparingly soluble in aqueous buffer, it could be dissolved in DMSO, methanol or TFE and then diluted into aqueous buffer for spectroscopic analysis. We chose to study the peptide pre-dissolved in DMSO by NMR, as this solvent is believed to have the smallest effect on peptide conformations. Despite some haze which is typical of partial aggregated samples, the soluble species of TDP-43(341-357) dissolved in 8% deuterated DMSO and 92% D_2O detected by NMR showed chemical shift values which are very similar to the longer peptides and to the standard "coil" values for short unstructured peptides (**Sup. Fig. 7**) and **Sup. Table 1G)**. For CD, it was necessary to study the samples pre-dissolved in methanol or TFE due to the high absorbance of DMSO in the far UV. The analysis of CD spectra of this peptide using the CDSSTR algorithm gave 6 % helix, 31 % β-sheet, 21% turn and 42% coil for TDP43(341-357) in 10% methanol and 2% helix, 41% β-sheet, 16% turn and 43% coil for TDP43(341-357) in 10% TFE. These results indicate populations of β-structure similar to those seen for aggregating WT TDP-43(342-366) (**Fig. 3 & Sup. Table 2**). Lastly, we found that TDP-43(341-357) samples strongly enhance ThT fluorescence, which suggest that its aggregates have amyloid-like characteristics. Taking these NMR, CD and fluorescence results together, we conclude that the TDP-43(341-357) segment can adopt an aggregating β-conformation. This conformation probably corresponds to the β-hairpin that is stable throughout the course of the MD simulation runs (*vide infra*) and could be the minimal aggregative segment of the protein *in vitro*.

Because these results raised several questions that could not be satisfactorily answered by these experimental approaches, we then decided to use Molecular Dynamics (MD) to gain further insight into the structure of this TDP-43 segment. The objective of these calculations was to characterize the putative structure and stabilizing interactions of the putative aggregative nucleus of TDP-43 formed by a monomer segment (residues 342-366) arranged in a β-sheet configuration. It is important to stress that these calculations neither provide information on the putative intramolecular interactions between this segment and other parts of the TDP-43 protein nor on the possible intermolecular interactions forming as the structural nucleus grows.

Computational Results

The TDP-43 segment 341-366 can adopt a well-defined anti-parallel sheet structure. The MD simulation of TDP-43(341-366) provides evidence that it adopts a set of preferred conformations. The structural deviation of a representative MD run is shown in **Fig. 4a**. The structure's RMSD is low (< 2 Å) and stable throughout the simulation, suggesting that this structure could be stable in physiological conditions. The residue-level secondary structure of TDP-43(341-366) during the simulation is shown in **Fig. 5a**. This analysis shows that the first and second β-strands, which are richer in Asn and Gln than the third strand, conserve their β-strand conformation up to 75 ns. Compared to the starting

model, the register of the strands is shifted one position so that residues Q343, Q344 and N345 of β-strand 1 pair with N352, N353 and Q354 of β-strand 2 (**Fig. 1a,c & 6a,b**). This set of six residues forming two β-strands seems to be the most stable element of structure, as it is conserved in almost all the conformations. Both strands can elongate to incorporate residues Q346 and S347 (β-strand 1) and S350 and G351 (β-strand 2). In addition, the third strand rapidly loses its canonical β-strand conformation to adopt a well-defined, but non-standard structure that interacts intimately with the second β-strand. We therefore think that this conformation is present at low population in the conditions of the spectroscopic experiments and may seed the formation of the β-sheet rich oligomeric structure by the WT peptide that is detected by our experimental CD analysis (**Fig. 3b**).

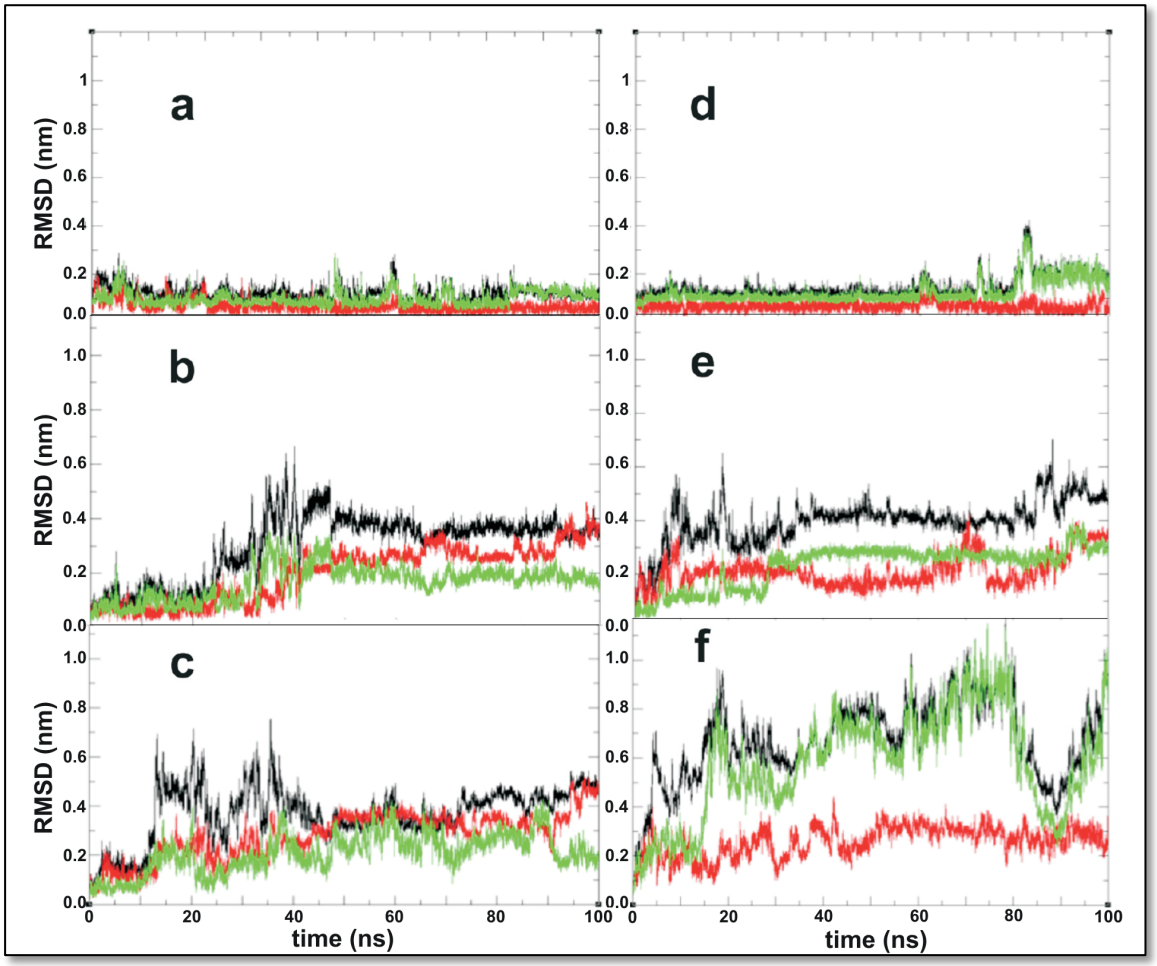


Figure 4: RMSD (nm) from the starting model for wt anti-parallel TDP(341-366).RMSD (nm, y-axis) from the starting model for wt anti-parallel TDP(341-366) versus time (ns, x-axis). **(a)** wt TDP-43(341-366), **(b)** Q343R, **(c)** N345K, **(d)** G348V, **(e)** N352S, **(f)** N352-N-Q-N-Q356 → A-A-A-A-A.
In **green**, residues 341 – 354, which correspond to the first two strands and their connecting loop; **red**, residues 355-366 which correspond to the second loop and the C-terminal residues. The overall average is shown in **black**.

To ensure a proper sampling of this structure, three additional 100 ns MD runs were performed with the all anti-parallel β-strand topology (**Fig. 1a**) in which each atom had a distinct starting velocity. A Principle Component Analysis and the observations that all three MD runs reach stable RMSD values over time and all successfully converge to the same fold (**Sup. Fig. 8**), strongly supports the validity of this structure. When three independent simulations were started from the alternative configuration (**Fig. 1b**), with the 3rd β-strand positioned in parallel to the 2nd, the 2nd and 3rd β-strands rapidly separate (**Sup. Fig. 9**). The 3rd β-strand then adopts a series of conformations, while the first two β-strands eventually separate and unfold in all three MD runs. Based on these observations, we conclude that the first (all anti-parallel) configuration obtained in **Fig.1a** is much more likely to be physiologically and pathologically relevant. For this reason, this all anti-parallel configuration was utilized to study the effects of the mutations.

In Gln343→ Arg, the Arg side chain promotes non-native interactions. MD simulations reveal large structural deviations (**Fig. 4b**) after 48 ns for the Q343R variant. The register of β-strands 1 and 2 is usually shifted relative to the WT structure. After only 63 ns, the β-hairpin structure unfolds (**Figs 1d & 5b**). The Q343 side chain projects into solvent in the model of WT TDP-43(341-366), suggesting that the loss of this side chain will not significantly affect this segment's conformation. However, the Arg's side chain that substitutes Q343 contains a guanidinum moiety that is closely related to guanidinum chloride, a potent chemical denaturing agent. Guanidinium chloride denatures proteins by interacting strongly with their aromatic and carbonyl groups. Structural snapshots reveal that after 60 ns, Arg343's guanidinium group frequently forms non-native H-bonds to carbonyl groups; some of which formed stabilizing interactions in the starting structural model (**Fig. 6c**). These aberrant interactions would tend to hold the segment in non β-hairpin conformations; in fact, the number of H-bonds formed by the Arg343 side chain increase notably after the native H-bond network breaks down (**Fig. 5b & 6d**). As a result, these aberrant interactions could account for why the spectroscopic experiments did not detect the formation of an oligomer rich in β-structure by the Q343R variant.

The variant Asn345→ Lys shows straighter β-strands with a better network of H-bonds. By MD simulation, β-strands 1 and 2 in the N345K variant were found to be longer on average than those of the WT sequence and more stable, as they do not unfold even after 100 ns (**Fig 5c**). These strands are also longer on average than they are in WT (**Fig. 5c**) and adopt the alternative strand register relative to WT that was also seen in the Q343R variant (**Fig. 1d**). This alternative strand register seems to cause this variant's RMSD values to be moderately large (**Fig. 4c**).

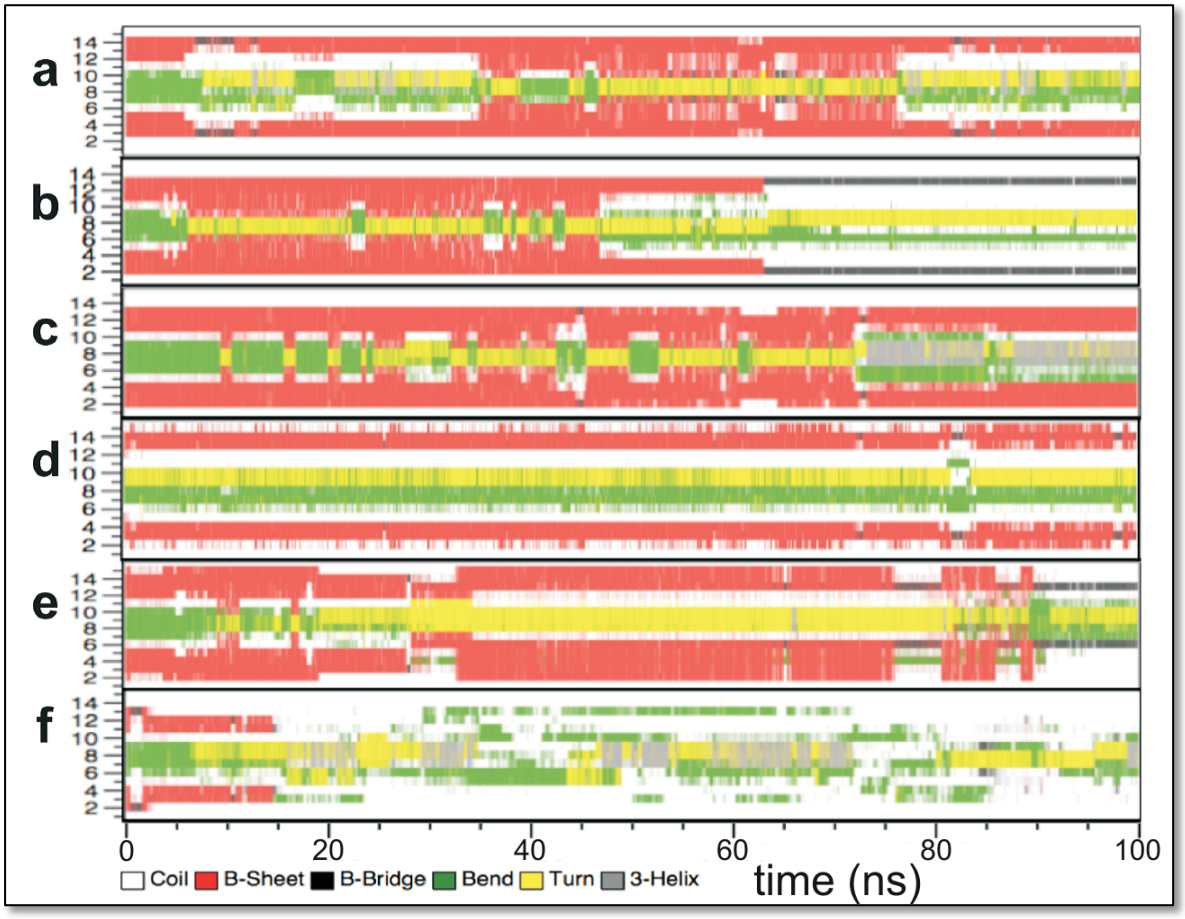


Figure 5. Secondary structure of TDP-43 residues 341-354 during simulation. Secondary structure of TDP-43 residues 341-354 during the 100 ns simulation, which correspond to the first two β-strands and their connecting loop and are numbered 1-16 as shown in **bluish-purple** in **Fig. 1**, over 100 ns of MD simulation (x-axis)**:** (**a**) wt TDP-43, (**b**) Q343R, (**c**) N345K, (**d**) G348V, (**e**) N352S and (**f**) N352-N-Q-N-Q356 → A-A-A-A-A.

In wild type TDP-43(341-366), the β-strands show a significant degree of twist. This distortion could be attributed, at least partially, to the formation of H-bonds between the Asn 345 side chain and the backbone amide groups of the adjacent residues (**Fig. 6e**). In the N345K mutant of TDP-43(342-366), however, the lysine side chain is unable to maintain these H-bonds and, as a consequence, straight β-strands with an increased number of H-bonds of improved geometry are observed **(Fig. 6f).** These structural changes could account for why the N345K variant's β-strand structure is expected to be longer and more stable. Considering the higher stability of the N345K variant's β-sheet, it seems puzzling that spectroscopic experiments showed that the WT peptide forms a β-sheet rich aggregate but the N345K variant does not. It is therefore possible that N345K's altered β-strand register and less twisted β-strand structure could be incompatible with oligomerization.

Mutating Gly348 → Val keeps β-strands 1 and 2 short and their connecting loop long. β-strands 1 and 2 in the WT TDP-43(341-366) can grow to encompass residues S342 - S347 and S350 - N355 (see **Fig. 5a** from ≈ 35 – 75 ns, for example) at the expense of their connecting loop. In fact, in the WT segment, this loop can be as short as just two residues: Gly348 and Pro 349. Whereas the effect of the Gly348→Val mutation on the overall

structural deviation is small (**Fig. 4d**), it is remarkable that the conformers with longer β-strand observed in WT are never seen along the trajectory of the G348V variant (**Fig. 5d**). Indeed, with Val at position 348, the β-strands remain short and their connecting loop is always long, as illustrated by the consistently large separation between the Cα's of Ser347 and Ser350 (**Fig. 6g,h).** This behavior, and some signs from NMR of altered flexibility in G348V, could be related to this variant's inability to form a multimer with increased β-structure in our spectroscopic experiments.

Asn352 → Ser favors an alternative β-hairpin structure. Many TDP-43 missense substitutions involve the creation of potentially novel phosphorylation sites. In our sequence of interest, such an occurrence could involve Asn352 and such an occurrence could have serious consequences on the structure. First of all, *even in the absence of phosphorylation, it has to be noted that* the N352S variant substitutes an Asn residue that actively participates in the structure of the second β-strand for Ser. During the course of the simulation and in the presence of this variant the β-strands 1 and 2 maintain a WT-like conformation for about 27 ns; afterwards this structure unfolds. It then refolds around 32 ns to an alternative conformation featuring a distinctive β-strand register (**Fig. 1e & 3e**) and a different β-hairpin structure (**Fig. 4e**) in which strand 1, composed of residues S342 – Q346 is two residues longer than β-strand 2, formed by residues N353-N355. It is notable that S352, the mutated residue, is excluded from the short β-strand. In fact, this residue is observed to form H-bonding interactions with the side chains of S347 and S350 and with the charged carboxylate group of E362 (**Fig. 6i,j**). These interactions may therefore account for this variant's singular β-strand register and loop conformation.

Phosphorylation of Ser 352 in the N352S variant strongly perturbs the structure. The effect of adding a phosphate group with a single negative change to Ser 352 was also studied (**Sup. Fig. 10**). The negative charge of the phosphate group, and its increased bulk, led to unfavorable interactions with the side chains of S347, S350 and E362. Along the course of the simulation, the β-strand structure adopted by residues 342 – 352 remains short and eventually is lost. The side chain of E362 turns outward and by the end of the simulation, it is positioned to form a salt bridge with the side chain of R361.

Multiple Ala substitutions are severely destabilizing. Substituting the residues of each of the three Q/N rich stretches in TDP-43(341-366) to Ala has been shown to decisively disrupt aggregation[11]. We therefore studied the behavior of the TDP-43(341-366) with residues Asn352-Asn-Gln-Asn-Gln356 substituted to Ala. As shown in **Fig. 4f & 5f**, it was observed that this variant rapidly deviates from the WT structure to adopt a pair of short β-strands composed of just two residues each; namely, Q343 & Q344 in strand 1 and G351 & N352 in strand 2. The strand register is like the one observed for the Q343R and N345K variants (**Fig. 1d**). This minimal β-hairpin unfolds after just 15 ns and the chain then adopts a broad variety of conformations.

Two 100 ns MD runs were also performed to study the effect of substituting P363, N364 and Q365 by Ala on the structure. In these simulations, the well-defined non-canonical structure breaks down and in one of the two runs, and the entire β-hairpin structure unfolds (**Sup. Fig. 11**). Based on these results and on the observed instability when the last segment is arranged in parallel to the second (**Sup. Fig. 9**) we propose that P363, N364 and Q365 also contribute to the overall conformational stability, at least at this first stage in the formation of the nucleus for aggregation. However, the ability of the TDP-43(341-357) peptide to adopt substantial populations of β-structure and to precipitate suggests that these stabilizing interactions are not essential for aggregation, at least *in vitro*.

3.4 Discussion

Pathological mutations in TDP-43 destabilize the β structure or shift the strand register in the 341-366 Gln/Asn-rich region. Amyloid structures have been implicated in many dementia and neurological disorders[29]. These structures consist of long fibrils composed of β-strands running perpendicular to the long fibril axis that are typically stabilized by efficient networks of H-bonds and by exceptionally tight chain packing. Recently, amyloids with out-of-register β-strands exposing unsatisfied H-bonding groups[30] or sporting exposed hydrophobic surface[31] have been proposed to be the most toxic and pathologically relevant forms. In this respect, therefore, it is interesting to point out that all the pathological mutations studied here: Q343R, N345K, G348V and N352S tend to be either structurally destabilizing or to shift the register of the β-strands of the 341-366Gln/Asn region so as to expose Asn and Gln residues. These changes, detected by MD, could be related to the inability of the NMR and CD experiments to detect the formation of β-sheet rich aggregates by the Q343R, N345K and G348V variants.

Of the approximately forty pathological mutations described to date in TDP-43, nine, including N352S, yield a new Ser or Thr residue, with many of these residues representing potential phosphorylation sites. Considering that the abnormal TDP-43 aggregates are hyper-phosphorylated in cells, it will be important to study the phosphorylation state of TDP-43 and their effects on its aggregation. In particular, we tentatively advance that eventual phosphorylation of the Ser hydroxyl group in the N352S variant breaks a network of H-bonds and as a consequence strongly perturbs the structure.

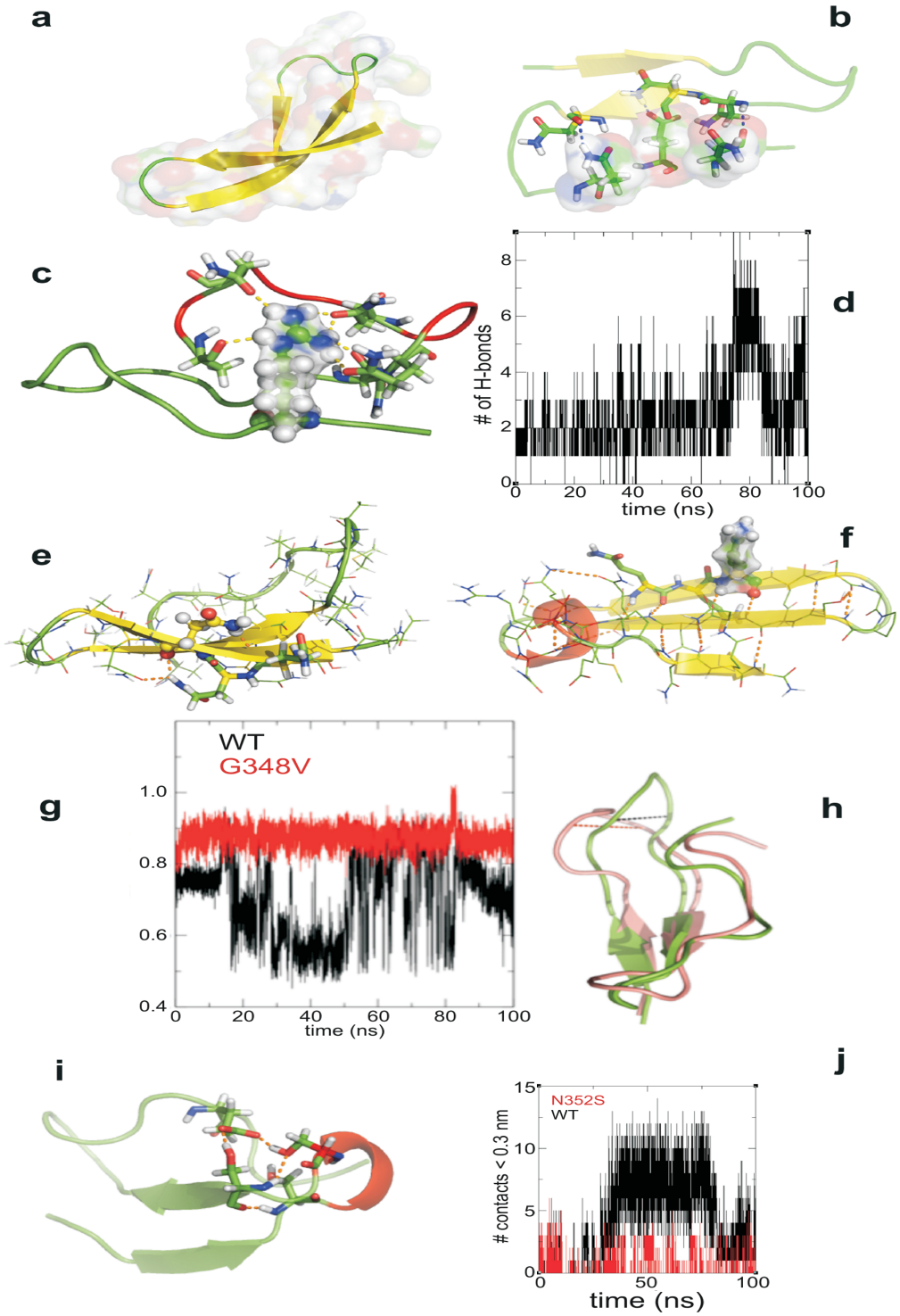


Figure 6. Representative 3D structures of TDP-43 (341-366) and its variants: **(a)**Starting structure of wt TDP-43 (341-366), showing the starting configuration. **(b)** wt TDP-43 (341-366), after 20 ns, showing changes in the β-strand register. **(c)** Snapshot taken along the course of the simulation of Q343R highlighting "non-native" contacts.**(d)** Number of H-bonds formed by residue 343: **black** = WT; **red** = Arg variant**(e)** Snapshot showing H-bonds formed by N345. **(f)** Mutation of N345 →Lys increases the number of H-bonds and the straightness of the β-sheet.**(g)** Loop 1 size as gauged by the distance between the Cα of S347 and S350; **black** = wt; **red** = G348V variant. **(h)** The size of loop 1 is always large in the G348→Val mutation.**(i)** Formation of contacts among the hydroxyl groups of S346, S350, S352 and E362 in the variant N352S.**(j)** The number of contacts (< 0.3 nm) among residues S346, S350, N352 and E362 in the WT segment (**black**) and the N352S variant (**red**).

TDP-43 deposits are commonly described as amorphous or granular aggregates and are generally found to be non-amyloid, since they neither bind amyloid specific dyes nor reveal amyloid fibrils when visualized by EM[6]. *In vitro*, of four peptides corresponding to segments from the C-terminal half of TDP-43, one formed amyloid-like fibrils, whereas the other three formed amorphous aggregates[32]. Another study of a series of peptides corresponding to the C-terminus of TDP-43 did not detect amyloid formation in the 341-366 region, although the peptides chosen; namely 337-349 and 350-362, would have separated the β-strand 1 from b-strand 2 in TDP-43 (341-366)[33], and we expect that this would have had a detrimental effect on β-sheet structural stability. Our ThT fluorescence results suggest that the β-sheet rich aggregates formed by the WT TDP-43(342-366) and TDP-43(341-357) peptide may be amyloid-like. Very recently, it has been shown that some forms of TDP-43 aggregates implicated in FTLD-TDP and ALS occasionally adopt amyloid-like structures in some conditions[34,35], although these findings' significance for TDP-43 pathology is presently unclear. Finally, the dichotomy between Asn-rich sequences, that tend to fold quickly into stable, less toxic amyloids, and Gln-rich sequences, that tend to form toxic aggregative intermediates and may not form stable amyloid[36], could be relevant for the TDP-43(341-366) segment, which is slightly Gln-rich (7 Gln vs. 6 Asn).

How does the rest of TDP-43 affect the 341-366 segment's behavior? Although previous results[17,37], and the spectroscopic and computational findings reporting here, indicate that TDP-43 segment 341-366 is sufficient to form a stable β-structure that could trigger aggregation, it is well known that other TDP-43 segments or external factors affect its tendency to aggregate. For example, presence of cleaved C-terminal fragments, N-terminal residues, the red/ox status of various Cys residues, and stress conditions can all contribute to trigger aggregation of this factor; for a recent review, see: ref. 2.

In particular, just to remain within the considerations made in this work that mostly concern missense mutations, a large number of pathology-linked mutations beside the four reported in this analysis have been mapped to the 54 residues preceding this segment and the 24 residues following it {for a most recent update, see ref. 38}. This extended region is rich in Gly residues that adopt unusual backbone angles and are good at making turns, two characteristics that are important in aggregate and amyloid formation in yeast prions[36]. Interestingly, a peptide corresponding to this extended segment of TDP-43 has also been reported to form amyloid fibrils[32]. Nevertheless, these amyloid fibrils do not enhance ThT fluorescence, which was suggested to be due the presence of a relatively high proportion of Gly residues. For a mechanistic explanation for the formation of these fibrils, the short side chain of Gly can constitute a "hole" that aids efficient side chain packing between β-strands within globular proteins, at protein – protein interaction interfaces and in amyloid[41]. Glys are also abundant in Gly-(Ala/Pro/Arg) dipeptide repeats aggregates that have been recently discovered and linked to FTLD/ALS neurological diseases[39,40].

The segment of TDP-43 composed of residues 341-366 is able to aggregate and drive the aggregation of full-length TDP-43 inside cells. Here, we provide spectroscopic evidence for the formation of a β-sheet rich oligomeric structure from a mostly disordered conformational ensemble by the WT peptide. Furthermore, MD evidence supports the observation that this segment forms a β-hairpin composed by residues 342 – 355 that is followed by a non-standard, although well-defined structural motif. As a result, these observations represent the first step aiming to provide a structural framework for interrogating the effects of pathogenic mutations and for formulating new hypotheses to guide future research in the aggregation process of TDP-43.

Finally, as highlighted here, protein chains may contain aggregation-prone or amyloidogenic segments that are normally "self-chaperoned" by other parts of the same protein chain[41] or by interactions with other partners. In these proteins, structured regions and binding to other factors can lock an aggregation-prone segment in a "safe" conformation. However, unfolding conditions or missense mutations can break the structured region's hold and unleash aggregative segments. Here, we show that the 341-366 residues in the C-terminus of TDP-43 may well represent one such element within this protein and suggest that more work to characterize the protein's structure at high resolution should be prioritized by future research trends. This will reveal not only the conformation of the TDP-43 341-366 segment in the context of the full length protein, but also interactions which could promote or hinder its aggregation. Such studies could provide further insight into how in-cell conditions or mutations increase TDP-43 aggregation and to design small effector molecules aimed at preventing, slowing down, or even reverting this process.

3.5 Methods

Peptides were synthesized and purified as previously described[17]. Their analysis by NMR and CD spectra corroborated their purity and identity. One additional peptide corresponding to residues 341 to 357, referred to here as TDP-43(341-357), was purchased from Genescript Corp. Its identity and purity (>95%) were confirmed by mass spectrometry, NMR and HPLC.

NMR Spectroscopy

1D ^{1}H and 2D ^{1}H TOCSY, NOESY and 2D ^{1}H-^{13}C HSQC and ^{1}H-^{15}N HSQC spectra were recorded on a Bruker 800 MHz AV NMR instrument which was equipped with a triple resonance (^{1}H,^{13}C,^{15}N) cryoprobe and z-axis gradients. The peptide concentration was generally 0.7 to 0.9 mM. All spectra were acquired in 10 mM K_2HPO_4 and 6 mM deuterated acetic acid (final pH ranged from 6.2 to 6.7). For more details on the NMR experiments, please see the **Supplementary Information**.

Circular Dichroism Spectroscopy

A JASCO J-710 spectropolarimeter was used to record far UV-CD spectra of the TDP-43 peptides. Spectra of the appropriate buffer or cosolvent were recorded at 5.0 and 25.0 ºC and were subtracted. Most samples were prepared by diluting small aliquots from the NMR samples into 10 mM K_2HPO_4 buffer (pH 8.2) or 3 mM KH_2PO_4/K_2HPO_4 buffer, pH 6.8. In an additional experiment, the WT lyophilized peptide was dissolved, a series of spectra were recorded at intervals of 30 to 60 minutes over several hours. Furthermore, two additional series of spectra were recorded on the TDP-43 peptide variants dissolved in the pH 6.8 buffer with 25% wt/v Ficoll, or alternatively, 0.15 M, 0.40 M and 1.00 M KCl. For a detailed description of the CD spectral analysis, please see the **Supplementary Information**.

ThT Fluorescence

A 1.0 mM stock solution of ThT (Sigma, St. Louis, USA) was prepared in 3 mM KH_2PO_4/K_2HPO_4 buffer, pH 6.8. Fluorescence samples contained 5 μL of ThT stock solution, whose final concentration was 50 μM and 25 μM of TDP-43 peptides. Amyloid $A\beta_{1-40}$ (rPeptide) and Ribonuclease A (Sigma type XII-A) were used as positive and negative controls, respectively. Spectra were recorded at 25.0 ºC on a Jobin-Yvon Fluoromax-4 instrument using 3 nm excitation and emission slit widths. The excitation wavelength was 440 nm and emission was recorded over 460 – 500 nm at a scan speed of 2 $nm \cdot s^{-1}$.

Molecular Dynamics Simulations

A starting model for TDP's aggregation promoting region was based on the known tendency of Asn, Ser and Gln residues to be located in β-strands and the high propensity of Gly and Pro for turn positions. The N-terminus and C-terminus were acetylated and amidated, respectively, to remove end charges and to more realistically mimic this segment's behavior in the context of a long protein. Two configurations, one with an all-

antiparallel orientation of the β-strands of the β-sheet (**Fig. 1a**) and a second with a mixed antiparallel/parallel β-strand orientation (**Fig. 1b**) were obtained and considered. The all atom models were built with the PyMOL package[42]. The force field utilized was AMBER99SB-ILDN[43]. MD was performed using GROMACS (version 4.5.5) software[44]. For more information on the Molecular Dynamics Procedures and Analysis, see the **Supplementary Information**.

3.6 Acknowledgments

We thank Dr. D. Pantoja-Uceda for assistance with NMR spectrometers and Prof. M. Bruix for critically reading the MS. The authors state no conflict of interests. This work was funded by grants CTQ 2010-21567-C02-02 to DVL from the Spanish Ministerio de Ciencia e Innovación and AriSLA (TARMA) to FEB and the Theirry-Latran Foundation (REHNPALS). The funders played no role in the experimental design or execution.

3.7 References

(1) E. Buratti, T. Dörk, E. Zuccato, F. Pagani, M. Romano, F. E. Baralle, EMBO J. 20 (2001) 1774-1784.
(2) E. Buratti, F. E. Baralle, Trends Biochem Sci. 37 (2012) 237-247.
(3) O. D. King, A. D. Gitler, J. Shorter, Brain Res. 1462 (2012) 61-80.
(4) M. Neumann, D. M. Sampathu, L. K. Kwong, Science 314 (2006) 130-133.
(5) T. Arai, M. Hasegawa, H. Akiyama, K. Ikeda, T. Nonaka, H. Mori, D. Mann, K. Tsuchiya, M. Yoshida, Y. Hashizume, T. Oda, Biochem Biophys Res Com 351 (2006) 602-611.
(6) M. Neumann, L. K. Kwong, D. M. Sampathu, J. Q. Trojanowski, V. M. Y. Lee, Arch. Neurol. 64 (2007) 1388-1394.
(7) T. F. Gendron, R. Rademakers, L. Petrucelli, J. Alzheimers Dis. 33 (2013) S35-S45.
(8) A. D'Ambrogio, E. Buratti, C. Stuani, C. Guarnaccia, M. Romano, Y. M. Ayala, F. E. Baralle, Nuc. Acids Res. 37 (2009) 4116-4126.
(9) R. A. Fuentealba, M. Udan, S. Bell, I. Wegorzewska, J. Shao, M. I. Diamond, C. C. Weihl, R. H. Baloh, J. Biol. Chem. 285 (2010) 26304-26314.
(10) H. Tsuiji, Y. Iguchi, A. Furuya, A. Kataoka, H. Hatsuta, N. Atsuta, F. Tanaka, Y. Hashizume, H. Akatsu, S. Murayama, G. Sobue, K. Yamanaka, EMBO Mol Med. (2013) 221-234.
(11) M. Budini, E. Buratti, C. Stuani, C. Guarnaccia, V. Romano, L. De Conti, F. E. Baralle, J. Biol. Chem. 287 (2012) 7512-7525.
(12) A. Yokoseki, A. Shiga, C. F. Tan, A. Tagawa, Ann. Neurol. 63 (2008) 538-542.
(13) N. J. Rutherford, Y. J. Zhang, M. Baker, J. M. Gass, PLoS Gene. 4 (2008) e1000193.
(14) J. Kirby, E. F. Goodall, W. Smith, J. R. Highley, Neurogenetics 11 (2010) 217-225.
(15) P. Kühnlein, A. D. Sperfeld, B. Vanmassenhove, V. van Deerlin, Arch. Neurol. 65 (2008) 1185-1189.
(16) M. Kamada, H. Maruyama, E. Tanaka, H. Morino, J. Neurol. Sci. 284 (2009) 69-71.
(17) M. Budini, V. Romano, E. Avendaño-Vázquez, S. Bembich, E. Buratti, F. E. Baralle, Brain Res. 1462 (2012) 139-150.

(18) S. Watanabe, K. Kaneko, K. Yamanaka, J. Biol. Chem. 288 (2013) 3641-3654.
(19) Y. J. Zhang, T. Caulfield, Y. F. Xu, T. F. Gendron, J. Hubbard, C. Stetler, H. Sasaguri, E. C. Whitelaw, S. Cai, W. C. Lee, L. Petrucelli, Hum. Mol. Gene. doi:10.1093/hmg/ddt116 (2013)
(20) A. Shodai, A. Ido, N. Fujiwara, T. Ayaki, T. Morimura, M. Oono, T. Uchida, R. Takahashi, H. Ito, M. Urushitani, PLoS One. 7 (2012) e52776.
(21) D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Skyes, J. Biomol. NMR 5 (1995) 67-81.
(22) M. V. Berjanskii, D. S. Wishart, J. Am. Chem. Soc. 127 (2005) 14970-14971.
(23) Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, J. Biomol. NMR 44 (2009) 213-233.
(24) S. Spera, A. Bax, J. Am. Chem. Soc. 113 (1991) 5490-5492.
(25) T. Crierpicki, J. Otlewski, J. Biomol. NMR 21 (2001) 249-261.
(26) K. L. Shaw, G. R. Grimsley, G. I. Yakovlev, A. A. Makarov, C. N. Pace, Protein Sci. 10 (2001) 1206-1215.
(27) M. Guo, P. Gorman, M, , M. Rico, A. Chakrabartty, D. V. Laurents, FEBS Letters 579 (2005) 3574-3578.
(28) H. X. Zhou, G. Rivas, A. P. Minton, Annu. Rev. Biophys. 37 (2008) 375-397.
(29) D. Eisenberg, M. Junker, Cell 148 (2012) 1188-1203.
(30) C. Liu, M. Zhao, L. Jiang, P. N. Cheng, J. H. Park, M. r. Saway, A. Pensalfini, D. Gou, A. J. Berk, C. Glabe, J. Nowick, D. Eisenberg, Proc. Natl. Acad. Sci. (USA) 109 (2012) 20913-8.
(31) R. Krishnan, J. L. Goodman, S. Mukhopadhyay, C. D. Pacheco, E. A. Lemke, A. A. Deniz, S. Lindquist, Proc. Natl. Acad. Sci. (USA) 109 (2012) 11172-77.
(32) A. K. H. Chen, R. Y. Y. Lin, E. Z. J. Hsieh, P. H. Pu, R. P. Y. Chen, T. Y. Liao, W. W. Chen, C. H. Wang, J. J. T. Tauang, J. Am. Chem. Soc. 132 (2010) 1186-1187.
(33) A. Saini, V. S. Chauhan, ChemBioChem 12 (2011) 2495-2501.
(34) J. L. Robinson, F. Geser, A. Stieber, M. Umoh, L. K. Kwong, V. M. vanDeerlin, V. M. Y. Lee, J. Q. Trojanowshi, Acta Neuropathol. 125 (2013) 121-131.
(35) E. H. Bigio, J. Y. Wu, H. X. Deng, E. N. Bit-Ivan, Q. Mao, R. Ganti, M. Peterson, N. Siddique, C. Ceula, T. Siddique, M. Mesulam, Acta Neuropathol. 125 (2013) 463-465.
(36) R. Halfmann, S. Alberti, R. Krishnan, N. Lyle, C. W. O'Donnell, O. D. King, B. Berger, R. V. Pappu, S. Lindquist, Mol. Cel. (2011) 72-84.
(37) M. Budini, E. Buratti, C. Stuani, C. Guarnaccia, V. Romano, L. De Conti, F. E. Baralle, J. Biol. Chem. 287 (2012) 7512-7525.
(38) S. Lattante, G. A. Rouleau, E. Kabashi, Hum. Mutation DOI:10.1002/humu.22319. (2013)
(39) P. E. A. Ash, K. F. Bieniek, T. F. Gendron, T. Caulfield, W. L. Lin, M. DeJesus-Hernandez, M. M. van Blitterswijk, K. Jansen-West, J. W. I. Paul, R. Rademakers, K. B. Boylan, D. W. Dickson, L. Petrucelli, Neuron 77 (2013) 639-646.
(40) K. Mori, S. M. Weng, T. Arzberger, S. May, K. Rentzsch, E. Kremmer, B. Schmid, H. A. Kretzschmar, M. cruts, C. Van Broeckhoben, C. Haass, D. Edbauer, Science 339 (2013) 1335-1338.
(41) P. K. Teng, N. J. Anderson, L. Goldschmidt, M. R. Sawaya, S. Sambashivan, D. Eisenberg, Prot. Sci. 21 (2012) 26-37.
(42) PyMOL, The PyMOL Molecular Graphics System, Schrödinger, LLC.
(43) K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, Proteins 78 (2010) 1950-1958.
(44) S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, Bioinformatics. 29 (2013) 845-854.

3.8 Supporting Information

3.8.1 Supporting Methods

A. Additional NMR Experimental Details

The samples contained 50 μM sodium 4,4-dimethyl-4-silapentane-1-sulfonate (DSS) as the ^{1}H chemical shift reference. ^{13}C and ^{15}N were referenced indirectly by multiplying by the gyromagnetic ratios[1]. Most spectra were recorded at 5 ºC. Additional 1D and 2D TOCSY spectra were at 15 ºC and 25 ºC to determine ^{1}HN temperature coefficients. Spectra were processed using Bruker Topspin software (v2.1) and assigned manually with the aid of the program Sparky[2]. The program TALOS+[3] was utilized to analyze the chemical shift data to determine backbone conformational preferences and dynamics. The solvent and NMR spectral parameters are summarized in the following table**:**

NMR Experiment	Solvent	Number of Scans	Number of Increments in 2nd Dimension	Mixing Time/Other Information
1D ^{1}H	90%H_2O/10%D_2O	32	--	--
TOSCY	90%H_2O/10%D_2O	24*	512	60 ms
NOESY	90%H_2O/10%D_2O†	64 - 80	512	150 ms
^{1}H-^{13}C HSQC (wt peptide)	99.9% D_2O	400	80	Transformed using linear prediction to 160 increments
^{1}H-^{13}C HSQC (variants)	99.9% D_2O	184	128	Transformed using linear prediction to 256 increments
TDP-43 (341-357)		400	128	
The sweep width in ^{13}C was 70 ppm and was centered at 39 ppm. The ^{1}H sweep width was 11 ppm in all spectra.				
^{1}H-^{15}N HSQC (Q343R)	90%H_2O/10%D_2O†	1024	128	Transformed using linear prediction to 256 increments
The sweep width in ^{15}N was 26 ppm and was centered at 118 ppm.				

* The number of scans for the TOCSY spectra of the G348V variant was 8 scans at 15 ºC and 16 scans at 25ºC. For TDP-43(341-357), 16 scans were recorded and this sample was dissolved in 99.9 % D2O.

† Additional NOESY spectra were recorded on the Q343R variant in 99.9% D_2O (# of scans = 32, # increments = 400, mixing time = 150 ms) and the G348V variant in 99.9% (# of scans = 64, # increments = 400, mixing times of 150 ms and of 250 ms) to look for $H\alpha_x - H\alpha_y$ NOES.

B. Circular Dichroism Spectroscopy Methodology and Analysis

CD Spectra were recorded using an instrument bandwidth was 1.2 nm and the scan speed was 50 nm·min^{-1}. Spectra were recorded at 5.0 ºC or 25.0 ºC in a 0.1 cm pathlength quartz cuvette over the wavelength range of 190 – 260 nm. The peptide concentrations were generally about 40 μM. Eight scans were recorded and averaged for each spectrum. Estimations of secondary structure content were performed using the algorithms SELCON3[4] and CDSSTR[5] as implemented on the webserver using ***Dichroweb***[6]: http://www.cryst.bbk.ac.uk/cdweb. Additional analyses with the K2D[7] and CONTINLL[8] algorithms were performed on the WT peptide spectra.

C. Molecular Modeling and Molecular Dynamics

To model the loops connecting the strands yielding the proposed hydrogen bonding patterns, dihedral angles were properly set with the GaussView v. 4.5. program[9]. The all atom models were built with the PyMOL package[10]. The force field utilized was AMBER99SB-ILDN[11]. MD was performed using GROMACS (version 4.5.5)[12] software. The cutoff distances for short-range interactions was 1 nm, and the PME algorithm[13] was used to describe long-range electrostatics. Dispersion-correction was applied.

The TDP-43 (341-366) polypeptide was placed in a 5-nm cubic box containing TIP3P water molecules[14] and energy-minimized using the steepest descent algorithm. Next, it was subjected to 500 ps of equilibration under an NVT ensemble to reach a temperature value of 300 K and a 2 ns equilibration under NPT conditions (1 bar), setting the initial conditions for the further MD run. In both cases, the Berendsen algorithm[15] was employed, with time constants of 0.1 and 2 ps, in each case. Three preliminary 50 ns simulations were first conducted on the TDP-43(342-366) polypeptide. In these runs and in subsequent simulations, the Nosé-Hoover thermostat[16,17] and the Parrinello-Rahman barostat[18] were employed, with time constants of 0.5 and 1 ps, respectively. A 2-fs time step was used by constraining all-atom bond lengths with the LINCS method[19]. For the initial equilibration periods, protein atom positions were restrained to allow solvent equilibration around the solute, whereas all the MD runs were produced in absence of any restraints. In all three 50 ns simulations, the first two β-strands were maintained but with an alternative strand register and the third β-strand adopted an uncommon, but well defined conformation (*as described in the* ***Results*** *section*). One of these simulations was selected; after re-initializing the starting trajectories, it was subjected to an additional 100 ns of MD. The last frame of this long simulation run was used as the starting point for modelling the point mutations.

MD simulations for wild type (WT) and each variant were carried out for 100 ns. To ensure that the WT structure was properly sampled, we performed a Principal Component (PC) Analysis to determine the cosine content of the first two PC's to ensure that there is not a lack of convergence of the trajectory. Additionally, three 100

ns simulations were performed, in which the last frame of the previous run was used as the starting point -as for the mutants- but with different initial velocities on each atom for every simulation, allowing the trajectory evolve through different ways to enhance sampling. Two independent simulations were run for Q343R, N345K and N352S each. MD structures were abstracted every 500 ps for visual examination and structural analysis. Their structural deviation from the starting model, as root-mean-square-deviations (RMSD), was calculated using the common analysis tools implemented in GROMACS, as well as hydrogen bonding patterns.

Approximately one hundred conformers per simulation run where examined visually with PyMOL. The program DSSP[20,21] was used to identify secondary structural elements. The WT TDP-43 structure was used as the basis for preparing the structures that include the missense mutations and the variants with the second or third strand's residues substituted for Alas.

3.8.2 Supporting Figures.

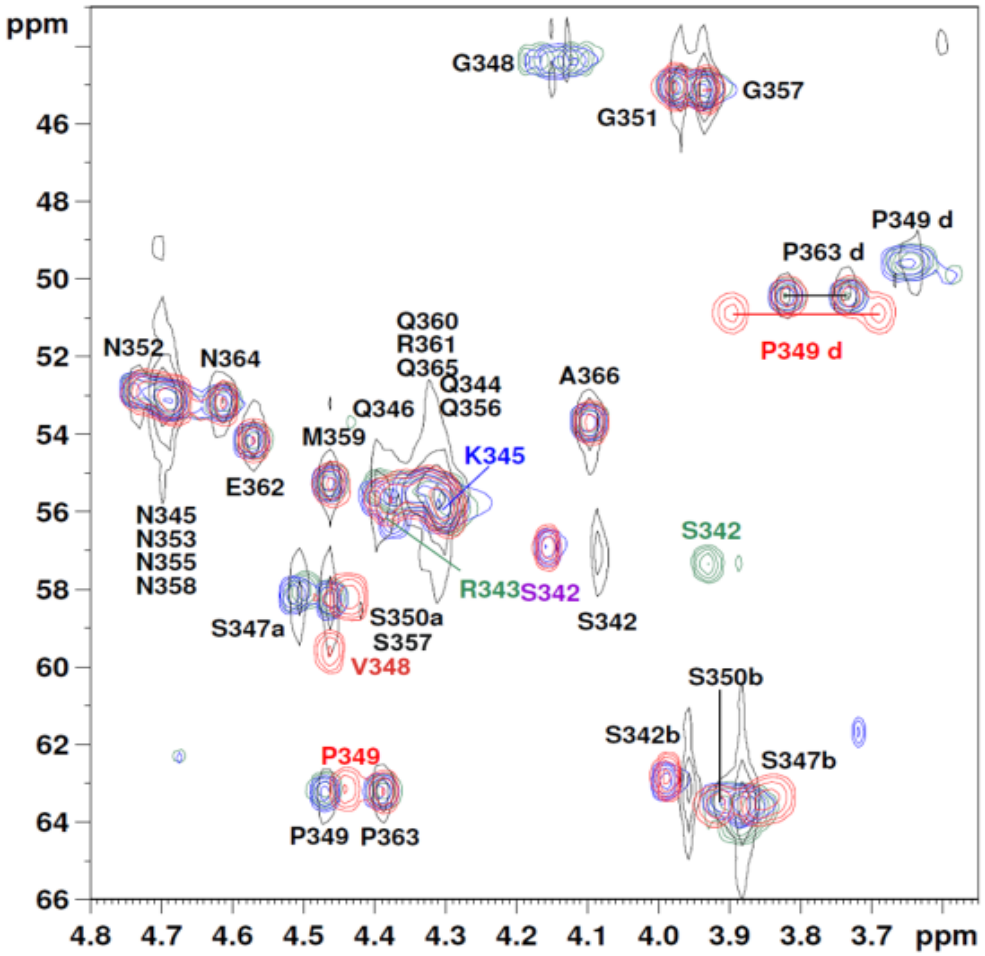


Figure S1: Comparison of the ^{1}H-^{13}C HSQC spectra of wt TDP43(342-366) and variants Q343R, N345K & G348V. ^{13}C-^{1}H HSQC spectra (alpha region) of wt TDP43 (342-366) in **black** and its variants Q343R (**blue**), N345K (**green**) and G348V (**red**). Most signals do not change substantially and are labeled in **black**. Some resonances that are unique to one spectrum are labeled with the color corresponding to their variant. The delta -CH_2- of P347 and P363 appear in the upper right hand corner of the graph. Note that the substitution of G348 by V causes to the 2 delta H of P347 to adopt distinct signals and also affects the Hα of S347. The signal : noise ratio and resolution in the wt spectrum is obviously worse than that of the other variants; this is due to the oligomerization of this peptide during the course of spectral acquisition.

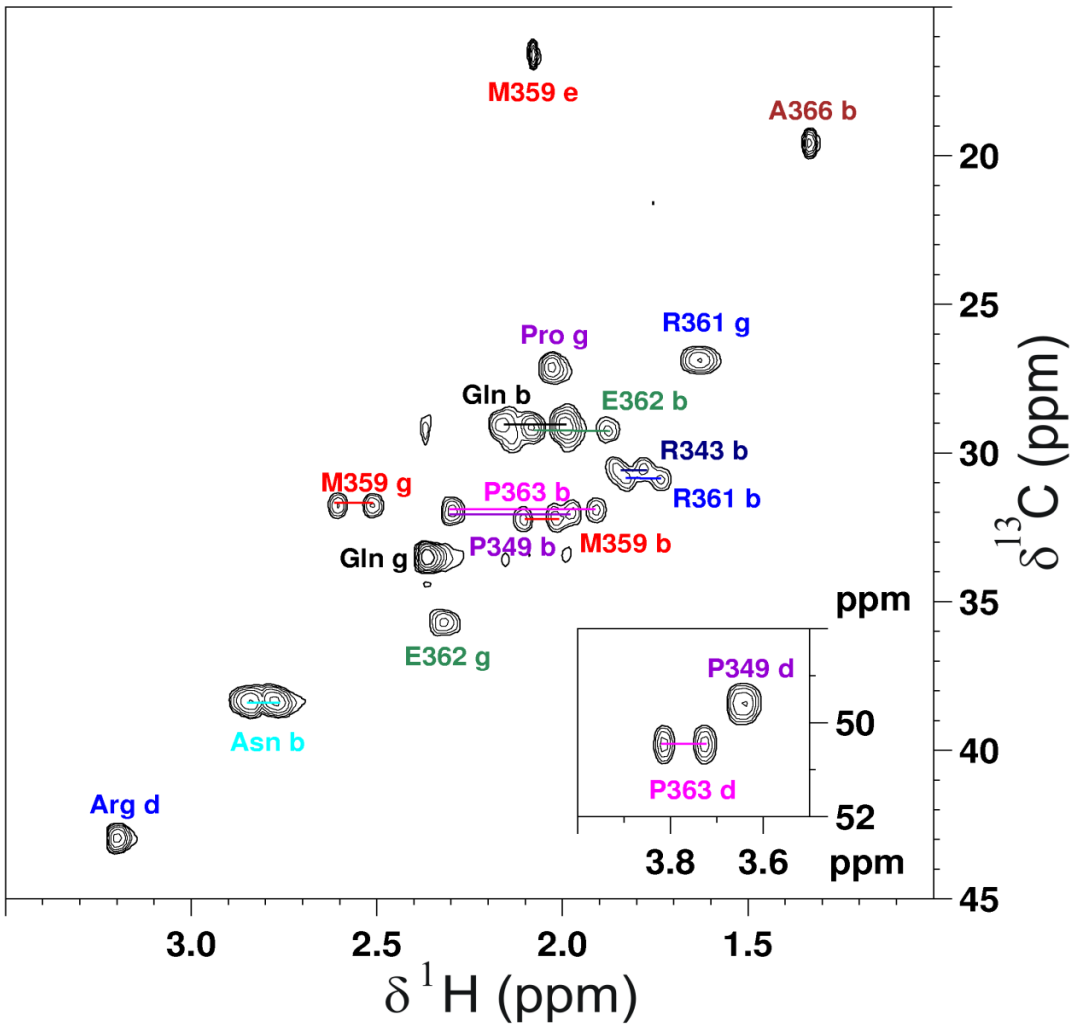


Figure S2: ^{1}H-^{13}C HSQC Spectrum (Side Chain Region) of TDP-43 (342-366) variant Q343R.The one letter amino acid code is used for signals assigned to a particular residue, and the three letter code is used when multiple signals arising from the same type of moiety from multiple residues are overlapped. The lower case letters "a", "b" *etc*. stand for "alpha" "beta", *etc*. Horizontal lines link geminal H which share the same C (*i.e*. γ -CH_2- M359). The oxidation state of Met 35 in the amyloid β peptide has been proposed to affect its propensity to aggregate[22]. Here, the chemical shift of the Met359 ε methyl group, is consistent with the sulfur being in the reduced state, (-CH_2-CH_2-S-CH_3) and not the sulfoxide state (-CH_2-CH_2-SO-CH_3). The chemical shift difference between the Pro ^{13}Cβ and ^{13}Cγ is 4.7 ppm. This indicates that both X-Pro peptide bonds are predominately in the *trans* conformation[23], as usually occurs for disordered peptides. The δ -CH_2- moiety of the prolines is shown in the **inset**.

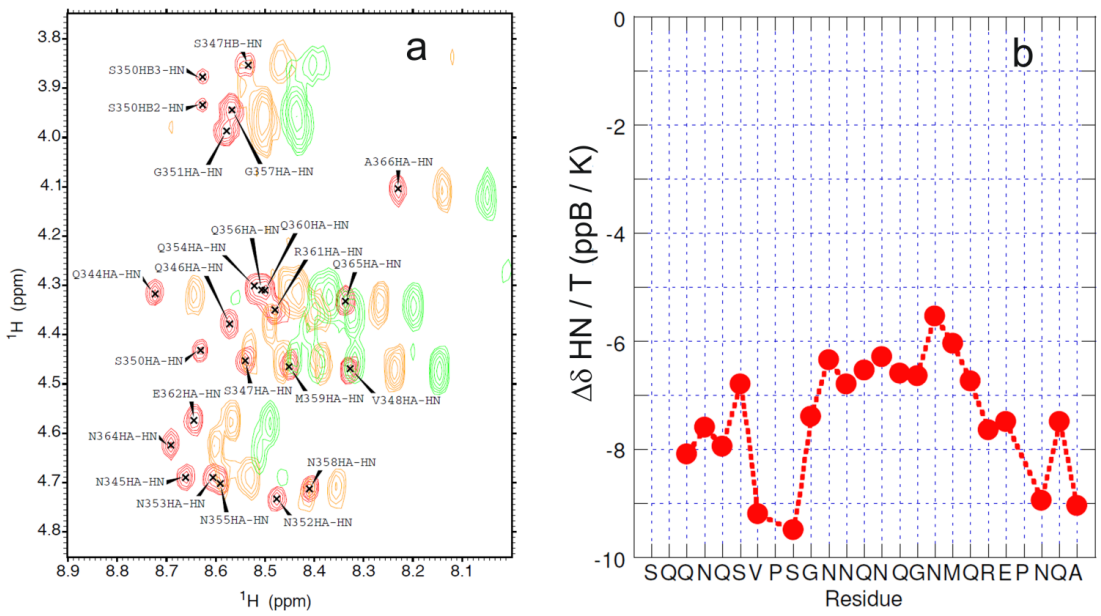


Figure S3: **a**. 2D ^{1}H TOCSY Spectra of TDP43(342-366) variant G348V at 5, 15 & 25 ºC and **b.** ^{1}HN temperature coefficients. **(a)** 2D ^{1}H TOCSY Spectra of TDP43(342-366) variant G348V at 5 ºC (red), 15 ºC (orange) & 25 ºC (green). ^{1}HN-^{1}Hα correlations at 5ºC are labeled. Note that the ^{1}HN signals move steadily downfield (to the right) with increasing temperature. **(b)** The change in the chemical shift of the ^{1}HN signals (in parts per billion) versus the change in temperature. Values more positive than -5 ppB/K indicate the HN groups form stable hydrogen bonds in secondary structure within polypeptides. The values reported here are all less that -5 ppB/K, which is consistent with the HNs being H-bonded to solvent water most of the time.

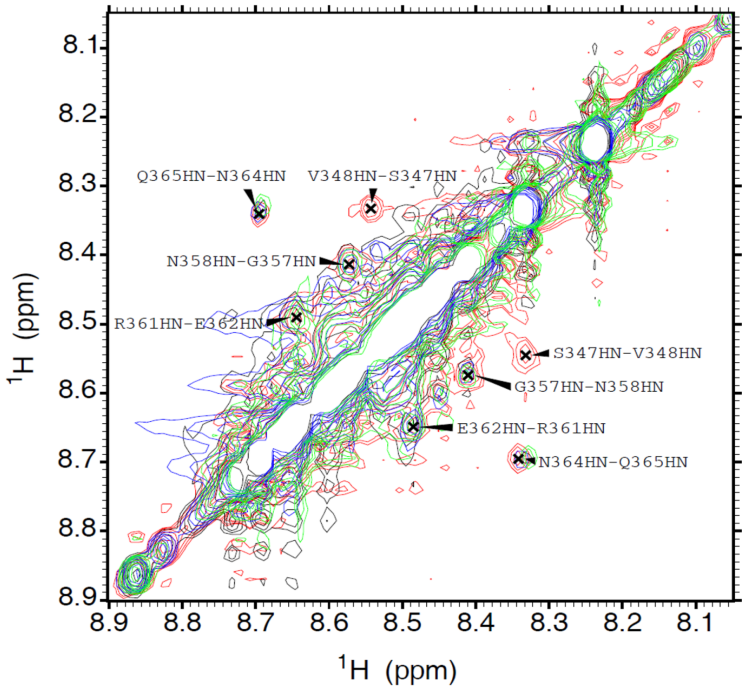


Figure S4: ^{1}HN NOEs of TDP43(342-366) wt and variants Q343R, N345K and G348V at 5ºC. Zoom of the ^{1}HN-^{1}HN region of the 2D NOESY spectra of TDP43(342-366) wt (**black**), Q343 (**blue**), N345K (**green**) and G348K (**red**). Mixing time = 150 ms. 1HN-1HN NOE crosspeaks are labeled.

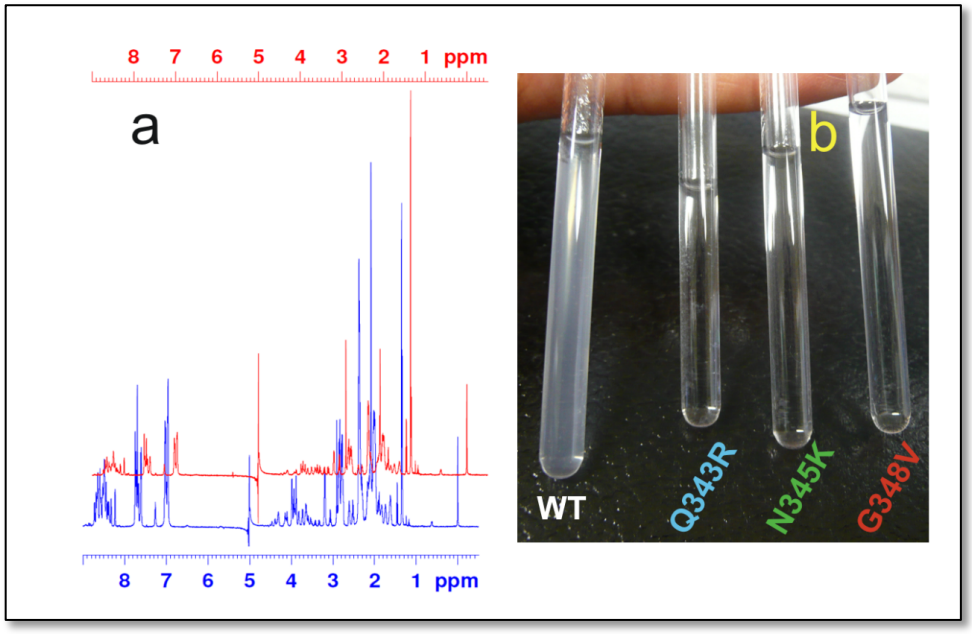


Figure S5: 1D ^{1}H NMR and Photographic Evidence for TDP43(342-366) WT aggregation. **(a)** 1D ^{1}H NMR spectrum of WT TDP43 (342-366) freshly dissolved (**blue**) and 16 h later (**red**). Note that the spectra are offset for clarity. Note that that the DSS reference signal at 0.00 ppm remains unchanged, but the signals corresponding to the peptide have lost 75% of their intensity after 16 h. **(b)** Photograph of the NMR sample of WT TDP43 (342-366), which is visibly hazy, and of Q343R, N345K and G348V variants, which are clear.

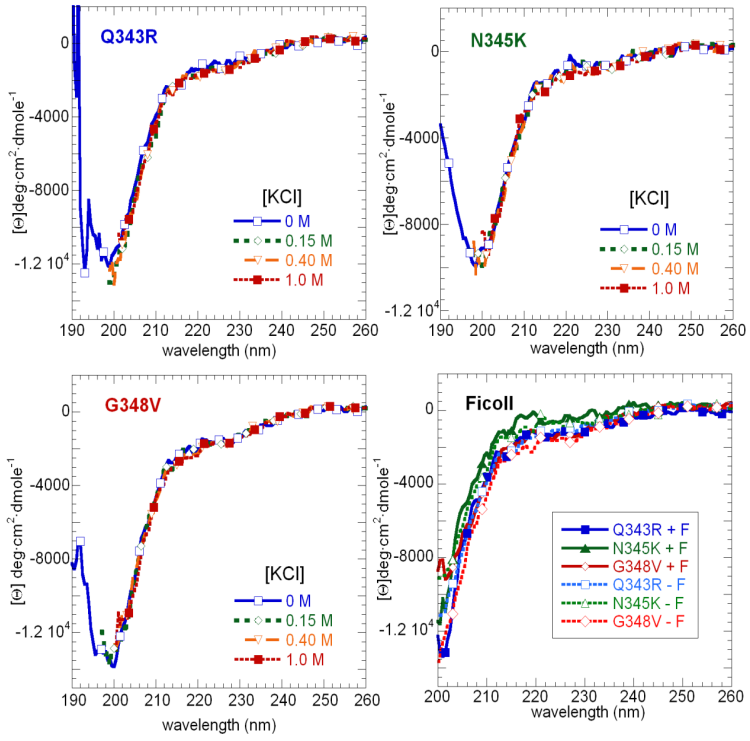


Figure S6: CD spectra of TDP43(342-366) variants Q343R, N345K & G348V in varying concentrations of KCl and 25 % wt/vol Ficoll. The measurements in KCl and Ficoll could not be extended to 190 nm due to their high far-UV absorbance. In the presence of KCl or Ficoll, none of the samples became visibly turbid and no significant changes in the CD spectra are evident.

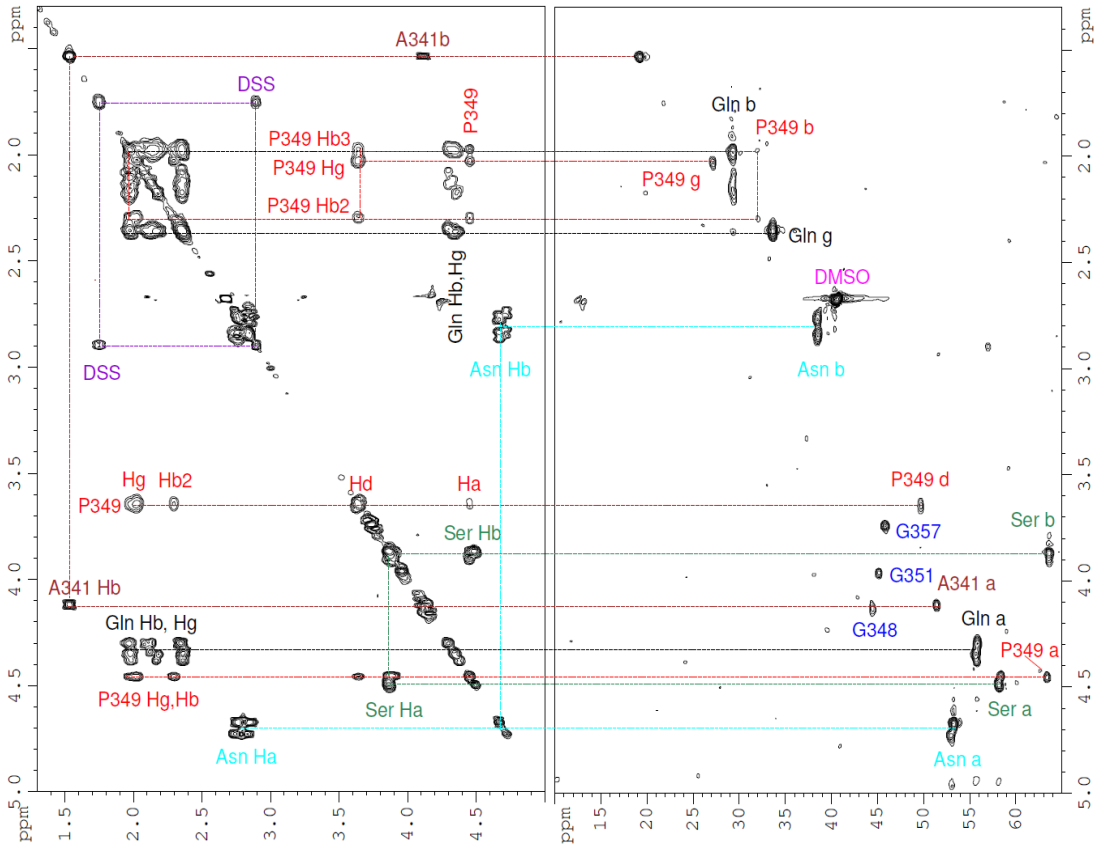


Figure S7: 2D ^{1}H TOCSY (left panel) and 2D ^{1}H-^{13}C HSQC (right panel) spectra in 8% DMSO (deuterated) and 92% D_2O at 5ºC of TDP-43(341-357). Intra-residual contacts are indicated by dotted lines. The one letter amino acid code is used for individual residues and the three letter code is used for groups of signals arising from the same type of residue. "DSS" (in purple) marks signals arising from the internal chemical shift reference compound. "DMSO" (magenta) marks the signal arising from the small amount of residual hydrogen in deuterated DMSO.

Time Interval (ns)	PC1 cosine content	PC2 cosine content
0-5	0.0176892	0.0153693
5-10	0.00495132	0.00300021
10-50	0.0743455	0.0554061
50-100	0.04856	0.0401999

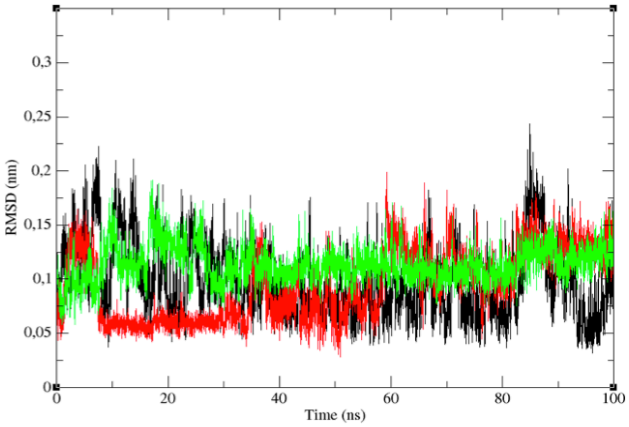


Figure S8: Additional Testing of the Convergence of the Wild Type Structure. The table (left) shows the cosine content values. These values were determined as followed: From the 100 ns run in the main text, we obtained the covariance matrix for the atomic positions by averaging the structure over the course of the simulation. Diagonalization of this matrix yielded the eigenvectors, regarded as the principal components (PCs) when the trajectory is projected onto them. Since the first two PCs contain the overall motion of this TDP segment due to its structural stability, the cosine content along different time frames of the trajectory was calculated. Low values were obtained in all of the cases, which indicate a lack of unidirectional motion. Instead, this means that the structure is well equilibrated and sufficiently sampled. As an additional measure of convergence, or stated more precisely, as an additional test of the lack of non-convergence, the RMSDs of three independent 100 ns MD runs that followed 2.5 ns of NPT equilibration are shown (right panel). The same structure was used in the three cases, which is the last frame of the initial 100ns simulation shown in the main text. For each run, the initial velocities on each atoms are different, in order to obtain a better sampling around that equilibrated structure allowing the trajectory evolve through different ways. The structure utilized for the comparative model for calculating the RMSD is the same one as was used for the Principal Component Analysis. Thus, we show that these three additional trajectories fluctuate around the averaged TDP-43 (341-366) structure from the first simulation (main text) with reasonably low RMSD values.

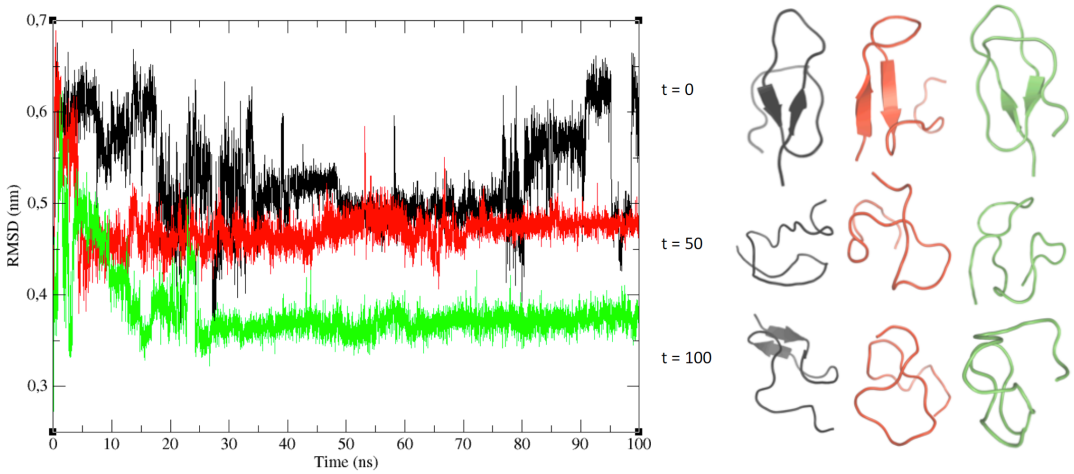


Figure S9. The β1 *antiparallel* β2 *parallel* β3 strand configuration for the TDP-43 (341-366) segment is unstable. Three simulations of the TDP-43 (341 – 366) segment initially arranged so that the 2nd and 3rd β strands are in parallel, as shown in **Fig. 1b** (main text). As in the other experiments, the structural models were first equilibrated with respect to temperature and pressure for 2.5 ns prior to the main 100 ns run. The RMSDs of the three simulations are shown in the *left panel*; the colors **black**, **green** and **red** represent the different simulations. Structure snapshots of the three simulations are shown on the *right panel*, at the start (top, t = 0 ns), middle (t = 50 ns) and end (t = 100 ns) of the main MD simulation. By the end of the initial 2.5 ns equilibriation period, which corresponds to t = 0 ns of the main MD run, structural alterations have already occurred, as reflected by RMSD values of 0.4 to 0.6 nm (*left panel*) and perturbations of the backbone structure (*right panels*). By 50 ns of simulation, the β-sheet structure has completely broken down. This result provides evidence that this topology is not the biologically relevant one.

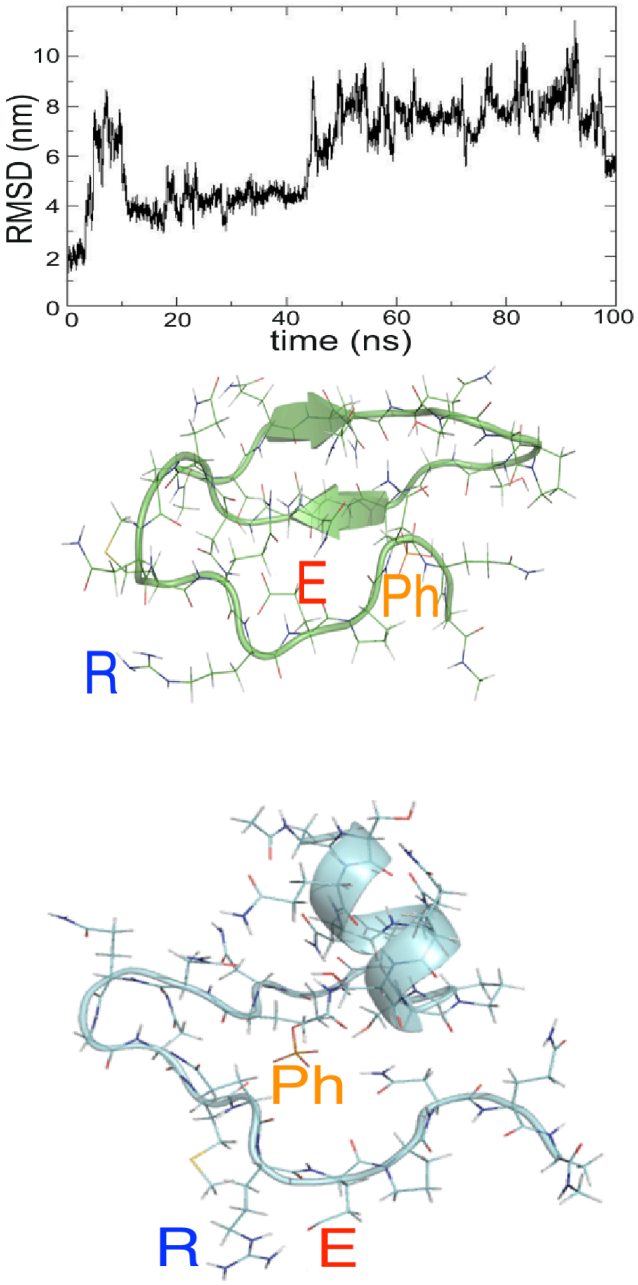


Figure S10. The phophorilation of S452 in the N352S variant strongly perturbs the structure of the TDP-43 341-366 segment. RMSD values for the simulation run are shown in the *upper* panel, and structural snapshots of the TDP-43 (341 - 366) segment at the start (**green**, time = 0 ns) and end (**cyan**, time = 100 ns) of one MD. The side chains of phosphoserine 352, R361 and E362 are labeled as **Ph**, **R** and **E**, respectively. It can be seen that the E362 side chain has already moved away from residue phosphoserine 352 by the end of the initial 2.5 ns equilibration period. During the course of the main 100 ns simulation run, the E362 side chain often pairs with that of R361 to form a salt bridge and the β-strands remain short. Eventually, the β-strands break down and the residues that composed the first strand occasionally form a short helix.

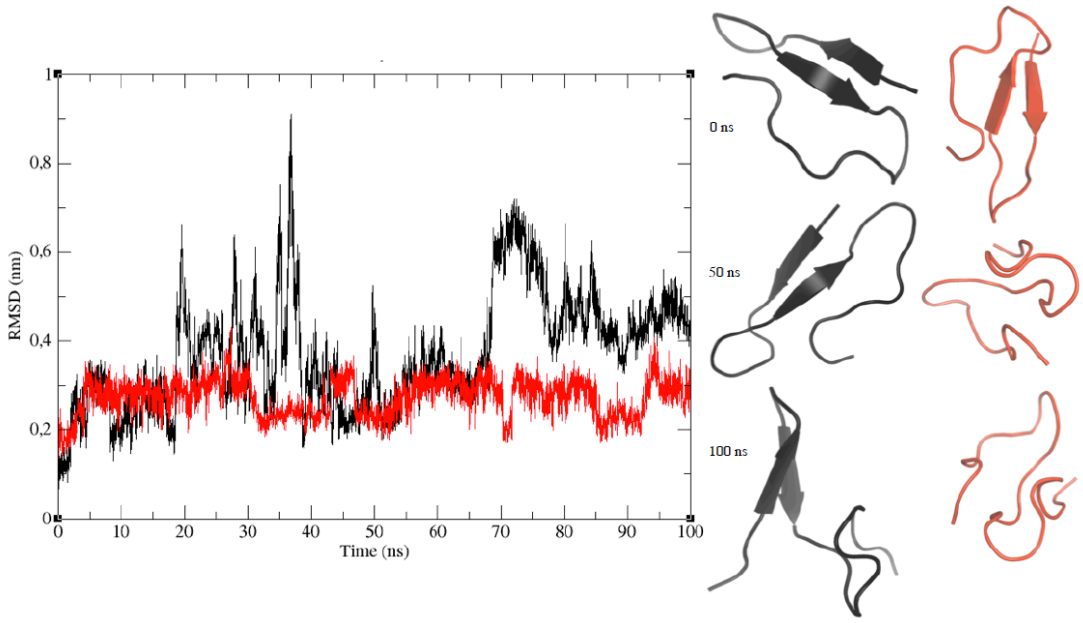


Figure S11. Mutation of P363, N364 and Q365 to Ala destabilizes the singular structure adopted by residues 359 – 366. Two 100 ns simulations were performed. By 50 ns, the secondary structure had broken down in one simulation run. By 100 ns, the contacts between the third segment with the first two strands were lost in both simulations. The observation of complete structure break down here in one out of two runs contrasts with the conservation of structure in all the runs with the wild type sequence. This provides evidence that P363, N364 and Q365 contribute to the conformational stability of the β-hairpin formed by the first sixteen residues of this polypeptide segment.

3.8.3 Supporting References

(1) Markley J. L., Bax A., Arata Y., Hilbers C. W., Kaptein R., Sykes B. D., . . . Wüthrich K. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. Pure Appl Chem 70, 117-142.
(2) Goddard T. D. and Kneller D. G. SPARKY. 3 edn., San Francisco
(3) Shen Y., Delaglio F., Cornilescu G. and Bax A. (2009) TALOS+, A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44, 213-233.
(4) Sreerema N., Venyaminov S. Y. and Woody R. W. (1999) Estimation of the number of helical and strand segments in proteins using CD spectroscopy. Prot Sci 8, 370-380.
(5) Johnson W. C. J. (1999) Analyzing protein circular dichroism spectra for accurante secondary structures. Prot Struct Funct Genet 35, 307-312.
(6) Whitmore L. and Wallace B. A. (2004) Dichroweb, an on-line server for protein secondary structure analyses from circular dichroism spectroscopic data. Nuc Acids Res 32, W668-W673.
(7) Andrade, M.A., Chacón, P., Merelo, J.J. and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot. Engineering* 6, 383-390.
(8) Van Stokkum I. H. M., Spoelder H. J. W., Bloemendal M., Van Grondelle R. and Groen F. C. A. (1990) Estimation of protein secondary structure and error analysis from CD spectra. . Anal Biochem 191, 110-118.
(9) Dennington R., Keith T. and Millam J. (2009) GaussView. 5 edn. Semichem Inc., Shawnee Mission, KS, USA
(10) PyMOL, The PyMOL Molecular Graphics System. 1.5 edn. Schrödinger, LLC.
(11) Lindorff-Larsen K., Piana S., Palmo K., Maragakis P., Klepeis J. L., Dror R. O. and Shaw D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950-1958.
(12) Pronk S., Páll S., Schulz R., Larsson P., Bjelkmar P., Apostolov R., . . . Lindahl E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845-854.
(13) Darden T., York D. and Pedersen L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 98, 10089-10092.
(14) Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W. and Klein M. L. (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79, 926.
(15) Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A. and Haak J. R. (1984) Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 81, 3684-3690.
(16) Nosé S. (1984) A unified formulation of the constant temperature molecular-dynamics methods. J Chem Phys 81, 511-519.
(17) Hoover W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A: At Mol Opt Phys 31, 1695-1697.
(18) Parrinello M. and Rahman A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52, 7182-7190.
(19) Hess B., Bekker H., Berendsen H. J. C. and Fraaije J. G. E. M. (1997) 3 LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18, 1463-1472.
(20) Kabsch W. and Sander C. (1983) Dictionary of protein secondary structure pattern recognition of hydrogen bonded and geometric features. Biopolymers 22, 2577- 2637.
(21) Joostern R. P., TeBeek T. A. H., Kneger E., Hekkelman M. L., Hooft R. W. W., Schneider R., . . . Vriend G. (2010) A series of PDB related databases for everyday needs. Nuc Acids Res doi:10.1093/nar/gkq1105.

(22) Hou L., Shao H., Zhang Y., Li H., Memon N. K., Neuhaus E. G., . . . Zagorski M. G. (2004) Solution NMR studies of the Aβ(1-40) and Aβ(1-42) peptides establish that Met 35 oxidation state affects the mechanism of amyloid formation 126: . J Am Chem Soc 126, 1992-2005.
(23) Schubert M., Labudde D., Oschkinat H. and Schmieder P. (2002) A software tool for the prediction of Xaa-Pro peptide bond conformation in proteins based on 13C chemical shift statistics. J Biomol NMR 24, 149-154.

Chapter 4

Structural Evidence of Amyloid Fibril Formation in the Putative Aggregation Domain of TDP-43

Logic will take you from A to B. Imagination will take you everywhere.
Albert Einstein

4.1 Abstract

TDP-43 can form pathological proteinaceous aggregates linked to ALS and FTLD. Within the putative aggregation domain, engineered repeats of residues 341-366 can recruit endogenous TDP-43 into aggregates inside cells. However, the nature of these aggregates is a debatable issue. Recently, we showed that a coil to β-hairpin transition in a short peptide corresponding to TDP-43 residues 341-357 enables oligomerization. Here, we provide definitive structural evidence for amyloid formation upon extensive characterization of TDP-43(341-357) via chromophore and antibody binding, electron microscopy (EM), solid-state NMR and X-ray diffraction. Based on these findings, structural models for TDP-43(341-357) oligomers were constructed, refined, verified, and analyzed using Docking, Molecular Dynamics and Semiempirical Quantum Mechanics methods. Interestingly, TDP-43(341-357) β-hairpins assemble into a novel parallel β-turn configuration showing cross-β spine, cooperative H-bonding and tight side chain packing. These results expand the amyloid foldome and could guide the development of future therapeutics to prevent this structural conversion.

4.2 Introduction

TDP-43 (transactive response DNA binding protein 43 kD) is a protein implicated in RNA regulation[1]. Its binding to nucleic acids takes place through two RNA recognition motifs (RRM) located at the N-terminus, where sequences for nuclear localization and export are also present. TDP-43 is also associated with neurodegenerative diseases, as aggregates of this protein have been found in patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration[2] (FTLD). TDP-43 may also play role(s) in the neurotoxicity of Aβ oligomers[3-5]. The nature of the TDP-43 aggregates and their pathological role(s) is still unclear. Although most TDP-43 aggregates do not resemble amyloid[2,6] some do under certain conditions[7,8]. For certain, short peptides corresponding to segments from other regions of TDP-43 can form amyloid-like fibrils *in vitro*[9-13] (Table S1). Moreover, the toxicity of the full-length protein has recently been attributed to the formation of noxious oligomers[14]. On the contrary, sequestering of all endogenous TDP-43 molecules into aggregates has also been suggested to cause a harmful loss of the native protein's function in its control of RNA metabolism[15]. From a pathway point of view, aggregate formation could be driven by several factors, that range from simple overexpression of this protein or the elimination of neighboring "self chaperoning" elements, as has been observed in RNase A[16]. For a review on the subject, interested readers are referred to Buratti and Baralle[1].

In summary, TDP-43 aggregates might: (i) be due to the elimination of neighboring "self chaperoning" elements, as has been observed in RNase A[16], (ii) play physiological roles in RNA and stress granules but eventually act as "TDP-43 sinks" to induce loss-of-function damage[15], or (iii) be neurotoxic through a gain-of-function mechanism. Most importantly, these scenarios do not necessarily exclude each other and may all be pathologically relevant. Therefore, understanding the dynamics of TDP-43 aggregation may represent a key step in the search and design of natural, synthetic or semi-synthetic compounds able to correct this process in patients' cells.

The localization of several pathological point mutations to the 341-366 segment suggests that it is relevant for disease processes[10]. The finding that a TDP-43 construct expressing multiple repeats of the C-terminal Q/N rich region spanning residues 341-366 is able to reproduce most of the pathological effects of the full-length protein aggregates in cells[17,15] constitutes strong evidence that this segment is relevant for formation of these aggregates. Recently, on the basis of experimental results from circular dichroism (CD) and nuclear magnetic resonance (NMR), we have characterized the structural transformation of a peptide corresponding to this segment in TDP-43, from a disordered monomer to an aggregating conformer rich in β-structure[18]. Using molecular dynamics (MD) simulations to test the stability of different plausible structural models, we showed that residues 341-357 formed a β-hairpin and residues 358-367 adopted a turn, followed by a well-defined yet non-canonical structure; however, how these monomers assemble and how the resulting oligomer's structure is formed remains unknown. Regarding this issue, the preliminary observation that a short

peptide corresponding to residues 341-357 aggregates more rapidly than 341-366 suggests that these residues are most crucial for self-association.

The main objective of this work is to characterize the structure of the oligomer formed by a peptide corresponding to residues 341-357 in TDP-43 using a variety of biophysical, biochemical, and computational methods. We also aim to rigorously test whether or not this aggregate has amyloid-like properties.

4.3 Results and Discussion

The short segment 341-357 of TDP-43 has been recently proposed to adopt a β-hairpin conformation[18]. Its sequence: AS**QQNQS**GPSG**NNQNQ**G, consists of two Gln/Asn rich regions (bold) separated by a turn (underlined). Because the peptide corresponding to this segment of TDP-43 protein aggregates quickly at concentrations appropriate for solution NMR, we studied a diluted sample by circular dichroism spectroscopy. The CD spectra indicated that β-structure is present in solution (Figure 1A and Table S2).

To gain additional structural information on the conformation of this peptide after aggregation, we used solid-state NMR. The 1D ^{13}C ssNMR spectrum of aggregated forms of TDP-43(341-357) shows significant resonance overlap, but a peak at 66.17 ppm can be unambiguously assigned to a Ser ^{13}Cβ (Figure 1B). The low field shifted value of this nucleus is consistent with the presence of β-sheet structure. To determine whether these β-rich aggregates are amyloid-like, we first performed Thioflavin T (ThT) and Congo Red (CR) binding assays, as amyloid-like fibers are well known to bind these two dyes. ThT fluorescence is strongly enhanced upon binding to amyloid fibrils, whereas CR increases its absorbance of visible light and experiences a red shift. As shown in Figure 1C and 1D, residues 341-357 of TDP-43 induce these spectral changes, which suggests that the aggregate behaves as an amyloid-like protein.

On the basis of these findings, we decided to determine the morphology of these aggregates with transmission electron microscopy (TEM). The fibrils were several hundred nanometers long and approximately 15-25 nm wide (Figure 1E). We observed flat, wide fibrils with a marked tendency to associate laterally. It is interesting that the morphology of these fibrils is very similar to that formed by expanded PolyQ (Q_{62}) segments[19] (Figure 1E), and distinct from the thinner (9 nm wide) nonbranching fibrils formed by the Aβ peptide[20].

The amyloid-like nature of these aggregates was further probed with two different conformational antibodies A11 and OC, which recognize oligomers and amyloid-like fibrils, respectively (Figure 1F). The A11 assay was positive, which strongly suggests the presence of TDP-43(341-357) oligomers similar to the small, diffusible multimeric species formed by several neurotoxic peptides[21], including polyQ[22] and Aβ[23]. In contrast, the binding of TDP-43(341-357) to the OC antibody

was negligible, in agreement with Wang *et al.*[24], who showed that the C-terminal motif of TDP-43 aggregates is not recognized by the OC antibody. This result is intriguing because the TEM images showed strong similarity with PolyQ fibers, which are recognized by the OC antibody[25] and thus may suggest a subtle structural difference between these fibrils.

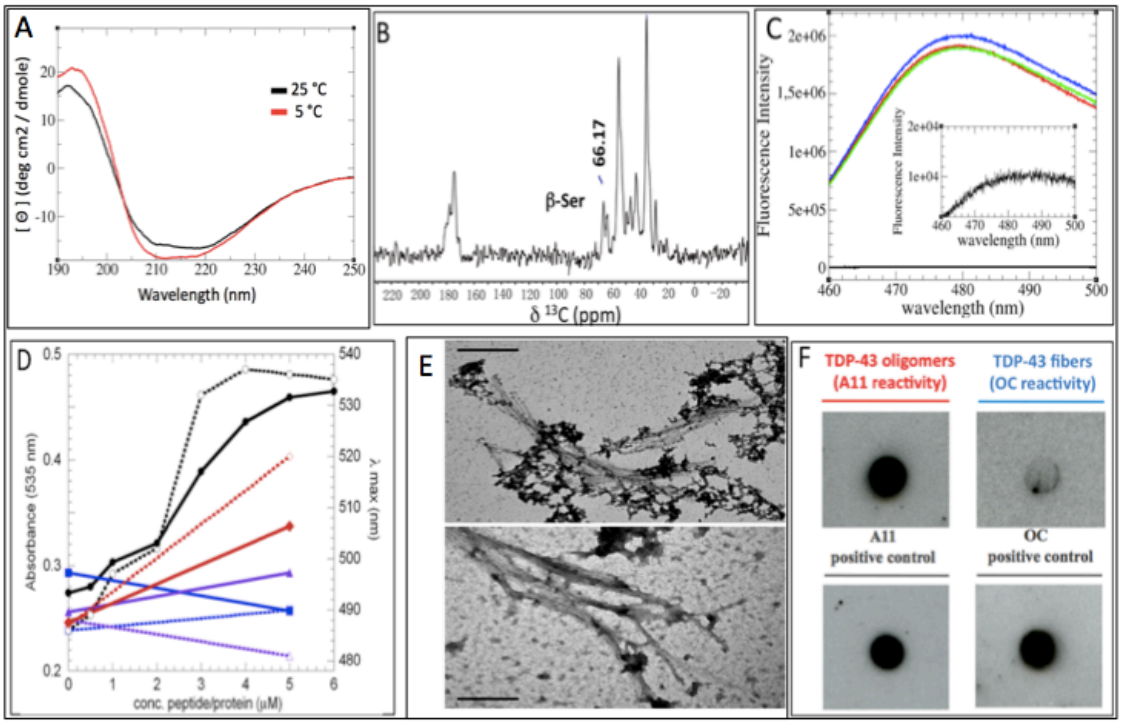


Figure 1. (A) Far-UV CD Spectra of a TDP-43 (341-357) at 25 ºC and 5 ºC (black and red curves, respectively), showing hallmarks of β-secondary structure in solution. (B) ^{13}C ssNMR spectrum; a peak arising from a Ser ^{13}Cβ atom in a β-conformation is indicated. (C) ThT fluorescence assay of three different samples (blue, red, and green curves) with a blank (black line) magnified, indicating the presence of amyloid structures. (D) Congo Red absorbance at 535 nm (solid symbols, solid lines, left y-axis) and absorbance maxima (open symbols, dotted lines, right y-axis) for in the presence of TDP-43 (341-357) (black circles), RNase A (blue squares), KIAβW (purple triangles) and $A\beta_{1\text{-}40}$ (red diamonds). (E) Representative TEM micrograph shows lateral association of TDP-43(341-357) aggregates (top) that resemble those formed by expanded PolyQ (Q_{62}) segments (bottom). The bar scale in both cases is 0.2 μm. (F) Antibody A11 recognizes TDP-43(341-357), which is consistent with the formation of smaller amyloid-like oligomers; in contrast, the OC antibody (specific for amyloid-like fibrils) shows weak/negligible binding. Prefibrillar oligomers and fibrillar species of $A\beta_{1\text{-}42}$ were used as positive controls for A11 and OC reactivity, respectively.

Considering our previous observations, therefore, these novel analyses have confirmed that the TDP-43(341-357) peptide evolves structurally, gaining β-structure over time and assembling into A11-reactive oligomers. Moreover, to test possible contributions from neighboring segments, we studied the conformation of TDP-43 (322-366) using NMR and found evidence for a random coil ensemble that aggregates

more slowly that TDP-43(341-357) (Figure S1); these results strongly suggest that flanking segments are not key for amyloid formation by TDP-43(341-357).

To gain further insight into this conversion, and make an atomistic model of the process, we carried out computational studies. Two TDP-43(341-357) monomers were placed in a simulation box and studied using three independent 100 ns MD runs. The three simulations started with different initial velocities on each atom. During all three runs, β-sheet formation occurs through the lateral association of the β-hairpins. Two different types of interactions were found, which we denoted as *in-register* (observed in two out of three runs) and *out-of-register* interactions (observed in one of three runs). The *in-register* β-strands permit each residue to interact with its counterpart in the strands above and below it, which we consider to be an optimal arrangement (Figure 2A). The *out-of-register* disposition involves an asymmetric packing of the residues within the strands, and did not seem to be a plausible intermediate towards formation of a mature fibril due to these less optimal interactions. Nevertheless, to corroborate the plausibility of both binding modes, we utilized Replica Exchange Molecular Dynamics (REMD) to improve sampling around these configurations. The results obtained provide additional evidence in favor of the *in-register* mode of β-hairpin association (Figure S2).

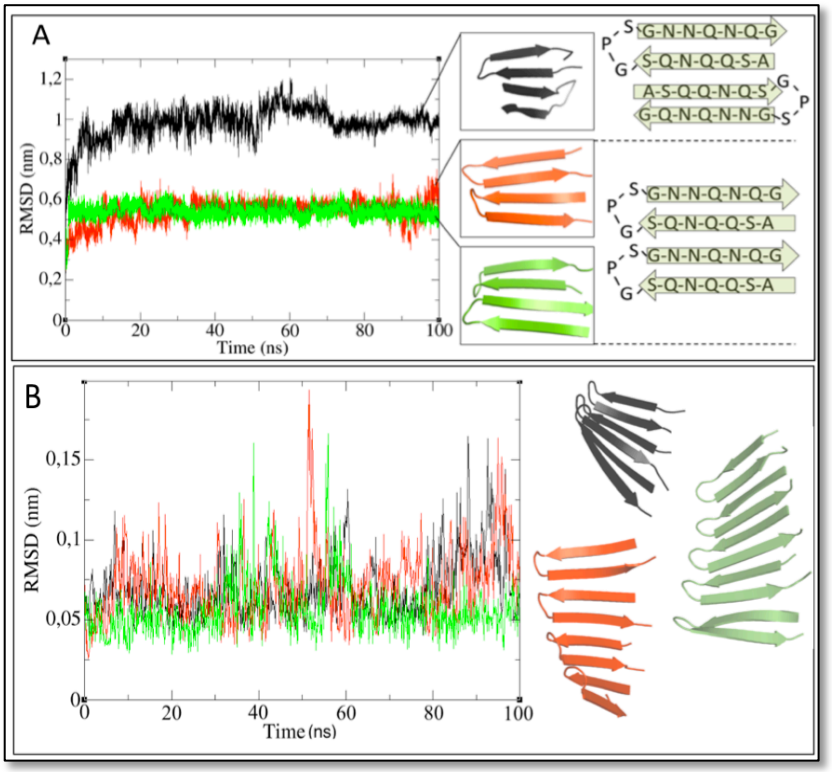


Figure 2. (A) RMSD corresponding to three independent MD simulations where lateral association occurs (Left). Note that the term *lateral* refers to the side by side packing of peptide backbones. (Left). Snapshots from the last frame of each of the structures are shown together with graphical representations of the *in-register* and *out-of-register* binding modes. The distance root-mean-square deviation (dRMSD) shows the same trend (Figures S3, S4, and S5). (B) dRMSD of the trimer (black), tetramer (red) and pentamer (green) fitted to their respective time-

averaged structures over all the frames. Snapshots at t=100 ns are represented for each system. All the RMSD calculations were performed with respect to the corresponding time-zero structures.

The in-register binding mode involves an extraordinary inter-residue complementarity because one strand of each β-hairpin is paired with the same residue above and below, which allows for optimal van der Waals interactions and the formation of an extensive network of hydrogen bonds throughout the Asn and Gln side chains. The stability of the interactions between these monomers is such that even upon heating *in silico* the system remained stable at temperatures up to 350 K. Interestingly, interactions of the strands within the β-hairpins were the first to dissociate, while the inter-strand interactions between different TDP-43(341-357) units were maintained. That is, the intermolecular interactions seem to be stronger than the intramolecular ones. Heating at 400 K corroborated this, where prior to dissociation the H-bonds were observed to break in the same order (Figure S6).

The quality of this structure was analyzed with PROCHECK[26], which showed that 91.7% of the residues are in the most favored regions, whereas the remaining 8.3% were in additional allowed regions of the Ramachandran map (Figure S7). This ensures the stereochemical quality of the system. On the basis of these results, we decided to dock additional monomers onto this pair of β-hairpins. Up to five monomers laterally associated were built this way and their structural stability was checked by MD simulations. After each 100 ns run, all three systems (trimer, tetramer, and pentamer) were found to be twisted and showed a remarkable stability (Figure 2B), as gauged by low dRMSD values. Supplementary Figure S8 provides further evidence of the structural stability of these systems, where the structures corresponding at times 0, 50, and 100 ns are superimposed to highlight the persistence both of twisting and the H-bond networks.

Amyloid-like oligomers typically display a remarkable resistance to chemical denaturants[27] and mechanical unfolding[28]. This exceptional conformational stability has been attributed to hyperpolarization of amide groups forming the H-bond network[29,30]. The elevated stability of the TDP-43(341-357) observed in the simulation runs mentioned above prompted us to test whether these structures are also able to exhibit the typical H-bonding cooperativity (HBC) present in amyloid-like fibrils. Semiempirical quantum mechanics methods have been proposed as an appropriate choice to study such interactions as they provide a good balance between computational cost and accuracy[31]. Therefore, we utilized the AM1 method to compute the energetics upon oligomerization through lateral association. The results of these calculations show that it is indeed more favorable to form systems of increasing size, which means that as more and more β-hairpins dock onto the growing aggregate, their H-bonds become stronger and stronger (Figure S9 and Table S3).

These results are able to reproduce both the H-bonding cooperativity and twisting that are characteristic of amyloid-like fibrils. Moreover, these findings are consistent with an aggregation pathway wherein the oligomer grows through lateral

association of monomers. Considering the spectroscopy-based evidence for β-conformation derived from CD and ssNMR experiments and given the positive results of the ThT and CR binding assays, the A11 immunoreactivity, and the morphology of the aggregates as revealed by TEM, which are in-line with amyloid-like aggregates, we tested whether these oligomeric β-sheets can pack to form a dry interface. To this end, we used HADDOCK[32] to perform fully flexible, solvated docking on two of TDP-43(341-357) pentamers. We obtained a cross-β spine structure, where the two β-sheets form a dry interface through precise side chain intermeshing with a high shape complementarity value[33] of 0.84. This value approaches that found (0.86) for the GNNQQNY heptapeptide of Sup35[34] and is significantly higher than typical Sc values for protease/protease inhibitors (0.71 to 0.76), subunit interfaces (0.70 to 0.74) or antibody/antigen contacts (0.64 to 0.68)[33]. A 100 ns MD simulation showed not only that this structure is stable, but also that the twist is preserved (Figure 3A).

Finally, the fiber diffraction of the peptide confirmed the cross-β association of the TDP-43 (341-357) monomers, which shows intra-β-sheet and inter-β-sheet reflections at 4.6 Å and 8.6 Å, respectively (Figure 3B). The latter value is smaller than the same reflection observed in most amyloid fibrils formed by other polypeptides, except for polyQ aggregates (see Table S4). These data establish the amyloid-like nature of the TDP-43(341-357) fibrils. X-Ray powder diffraction of the fibers that were used for the ssNMR experiments showed the same two major reflections (Figure S10), with d-spacing corresponding to the above distances, which makes this motif unique with respect to previously reported amyloid structures. This result can reconcile all our experimental findings with the predicted aggregative pathway based on β-hairpin formation and lateral association.

Furthermore, inspecting the dry interface (Figure 3C) of this aggregative motif clarifies why the pathological substitutions Q343R and N345K do not aggregate *in vitro*[18]. In both cases, the length of the mutated side chains will not allow for the proper side chain intermeshing, as can be inferred from Figure 3D and 3E, because they are too long for maintain the two mating β-sheets at the distance of 8.6 Å observed here and in polyQ fibrils[35] or even the longer distance of 10 Å seen in most amyloid fibrils (Table S4).

4.4 Conclusions

On the basis of all the experimental and computational results, we conclude that TDP-43(341-357) forms an amyloid-like fibril with a novel β-turn topology in which *each β-hairpin associates in parallel to contribute two β-strands to the same β-sheet*. It is interesting that in this structural model TDP-43(341-357) shows a "β-turn" configuration, which has been recently reported for polyQs[36] and in contrast with the "β-arc" topology observed for Aβ in which β-hairpins contribute one β-strand to two different β-sheets. This β-arc fold is not likely to occur in TDP-43(341-357) as it would imply breakage of the inter-strand H-bonds favored by the Gly-Pro turn. In fact, it was

not observed in our REMD and MD simulations (See Figure S11). Nonetheless, TDP-43 and polyQ aggregates share a similar morphology, as seen by EM, and β-turn topology. However the configuration of consecutive hairpins is parallel in the case of TDP-43(341-357) and anti-parallel for polyQ aggregates.

In conclusion, we have shown that the TDP-43(341-357) peptide can form aggregates that are recognized by the conformation-specific A11 antibody, and fibrils with amyloid-like dye-binding characteristics. In contrast to most amyloid-like fibrils, but akin to those formed by polyglutamine, TDP-43(341-357) fibrils show a rather unusual flat morphology, as seen by EM, and short inter β-sheet separation, as measured by X-ray diffraction. Finally, we advance a mechanism for oligomerization and an atomic resolution structural model for the TDP-43(341-357) amyloid-like fibril in which β-hairpins combine in a novel parallel β-turn configuration. These results advance our understanding of aggregates in TDP-43 pathologies and could aid the development of therapeutics specifically aimed at slowing down or inhibiting this process.

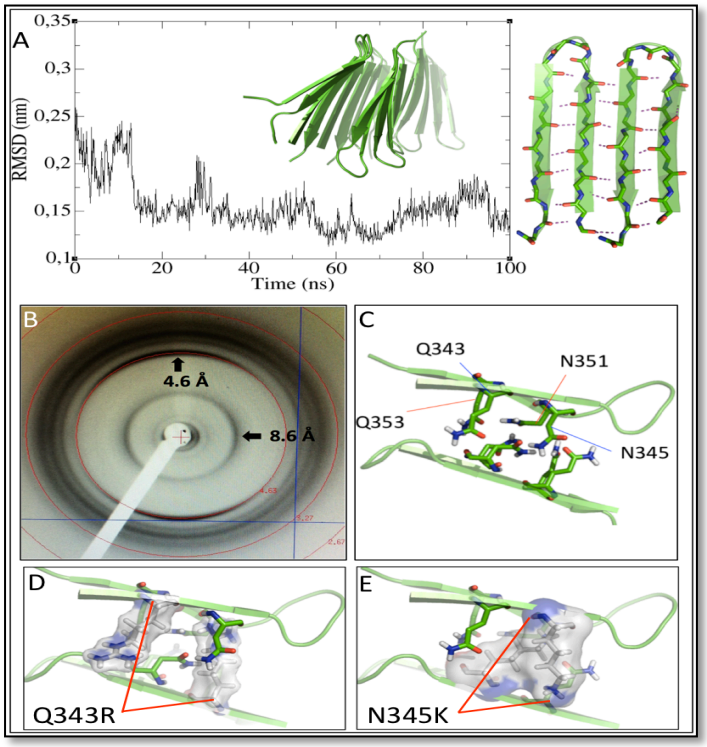


Figure 3. (A) RMSD of two TDP-43(341-357) pentamers packed forming a cross β-spine from a 100 ns MD simulation. Each pentamer is a ten-stranded β-sheet formed by five anti-parallel β-hairpins, whose stability is evidenced by low RMSD values during MD simulations, calculated with respect to the time-zero structure. (B) X-Ray diffraction pattern collected from fibers confirming the presence of a cross β-spine. The reflection at 4.6 Å corresponds to the inter-strand distance, whereas that at 8.6 Å is the inter-sheet distance. (C) Gln and Asn sidechains contact and form a dry interface between the two β-sheets, with a high shape complementarity value of 0.84. The Q343R (D) and N345K (E) substitutions are incompatible with this supramolecular organization, and in consequence, peptides carrying these substitutions do not oligomerize[18].

4.5 Methods

4.5.1 Experimental methods

A peptide whose sequence: ASQQNQSGPSGNNQNQG corresponds to TDP-43(341-357) was purchased from Genescript (Piscataway Township, NJ). Its purity was over 95% as assessed by HPLC. NMR spectroscopy and mass spectrometry confirmed the peptide's composition. The peptide's concentration in solution was determined by weight and by far-UV absorbance, using ε_{205nm}= 48080 $cm^{-1} \cdot M^{-1}$ at 205 nm calculated using the sequence and the parameters of Anthis & Clore[37].
PolyQ-containing protein was cloned by PCR using a plasmid containing 62 Qs as template, kindly provided by Dr. Yoshitaka Nagai[38]. The PCR insert was cloned into the pET28a vector (Novagen) using the NheI and XhoI restriction sites. With this cloning strategy, the MGSSHHHHHHSSGLVPRGSHMAS amino acid sequence remained at the N-terminus of the protein. Expression and purification of the protein is fully described elsewhere[28].

Circular Dichroism Spectroscopy. Far UV-CD spectra were recorded using a JASCO J-710 spectropolarimeter. The instrument bandwidth was 1.2 nm and the scan speed was 20 $nm \cdot min^{-1}$. Spectra were recorded over the wavelength range of 190-260 nm at 25.0 ºC using a 0.1 cm pathlength, with a peptide concentration of 105 μM in 3 mM KH_2PO_4/K_2HPO_4 buffer (pH 6.8) at 5 and 25 ºC. Eight scans were recorded and averaged for each spectrum, and the buffer reference spectrum was subtracted. For a detailed description of the CD spectral analysis, we refer the reader to the SI Methods**.**

Solid-State NMR Spectroscopy. The TDP-43(341-357) peptide (25 mg) was dissolved in 10 mM KH_2PO_4/K_2HPO_4 (pH 6.3) and allowed to form fibrils during 7 days. Following solvent removal, the fibrils were packed in a 4 mm Varian rotor and measured on a Chemagnetics Infinity 300 MHz instrument in H/C double resonance mode with an APEX double resonance Magic Angle Spinning (MAS) probe. Cross Polarization (CP) experiments were performed with a Hahn-Echo to suppress broad background signals. TPPM decoupling (80 kHz) was applied during acquisition.

Thioflavin T Fluorescence. Following the procedure in LeVine[39], a 1.0 mM stock solution of ThT (Sigma, St. Louis, MO) was prepared in 3mM KH_2PO_4/K_2HPO_4 buffer, pH 6.8. Fluorescence samples contained 5 μL of ThT stock solution, whose final concentration was 50 μM, with 25 μM of TDP-43 peptides. Amyloid $A\beta_{1-40}$ (rPeptide) and Ribonuclease A (Sigma type XII-A) were used as positive and negative controls, respectively. Spectra were recorded at 25 ºC on a Jobin-Yvon Fluoromax-4 instrument using 3 nm excitation and emission slit widths. The excitation wavelength was 440 nm and emission was recorded over 460-500 nm at a scan speed of 2 $nm \cdot s^{-1}$.

Congo Red Binding. Congo Red spectroscopic assays were performed as described by Klunk *et al*.[40]; 5 μM of CR stock solution (Sigma) were prepared in 5 mM KH_2PO_4 and 150 mM NaCl, and used to test binding to TDP-43(341-357) at concentrations ranging

from 1 to 5 μM. Amyloid $A\beta_{1-40}$ (rPeptide) was used as a positive control, whereas Ribonuclease A (Sigma type XII-A) and KIAβ, a non-aggregating β-hairpin peptide[41], were used as negative controls. Congo Red absorption spectra were recorded from 440 to 600 nm in a dual beam Cary 210 UV-vis spectrometer at room temperature, at a ½ $nm \cdot s^{-1}$ scan speed. The samples were aged for six days to ensure aggregate formation and study their possible amyloid-like nature.

Transmission Electron Microscopy. TDP-43 peptide samples were incubated for ~14 days at 37 ºC at a concentration of 50 μM in PBS [pH 7.0] in the presence of 5% DMSO and with no stirring. PolyQ samples were incubated for 20-30 days at 37 ºC at a concentration of 10-40 μM in PBS [pH 7.4] in the presence of 0.02% NaN_3 and with no stirring. For fibrillogenesis assay, samples diluted to concentrations of 5-50 μM (10 μl) were adsorbed onto carbon-coated 300-mesh copper grids (Ted Pella) and negatively stained for 60 s using 1-2% uranyl acetate. Immediately before use, the carbon-coated grids were glow-discharged to enhance their hydrophilicity using an Emitech K100X apparatus (Quorum Technologies). The formation of amyloid fibrils was analyzed on a JEOL 1200EX II (Jeol Limited) electron microscope equipped with a CCD Megaview III camera (Olympus Soft Imaging). The images were acquired at a magnification of 120.000 x and a voltage of 80 kV.

Dot Blot Analysis. TDP-43 samples were incubated for approximately 2 weeks at 37 ºC at a concentration of 50 μM in PBS with 5% DMSO, pH 7.0 and without stirring. For Dot Blot analysis, 2 μl of each TDP-43 peptide sample was spotted onto a nitrocellulose membrane. After blocking for 1 h at RT with 10% non-fat milk in Tris-buffered saline (TBS) containing 0.01% Tween 20 (TBS-T), the membrane was incubated with the polyclonal specific antioligomer A11 antibody specific for neurotoxic oligomers[21] (Life Technologies) or the monoclonal antibody OC (Millipore) which recognize mature amyloid-like fibrils,[25] diluted 1:1000 in 3% BSA TBS-T, at RT for 1 h. The membranes were washed three times for 5 min each with TBS-T before incubating with anti-rabbit HRP conjugated anti-rabbit IgG (GE Healthcare) diluted 1:5000 in 3% BSA/ TBS-T at room temperature for 1 hour. After washing the membranes thrice in TBS-T buffer, the blots were developed with ECL Plus chemiluminescence kit from Amersham-Pharmacia (GE Healthcare). Prefibrillar oligomers and fibrillar species of $A\beta_{1-42}$, served as positive controls for A11 and OC reactivity, respectively.

X-Ray Fiber Diffraction. In an Eppendorf tube, 2 mg of peptide powder were dissolved in 0.8 mL of water. The solution was vortexed at room temperature. The resulting cloudy solution was allowed to settle for 7 days to form fibers, after which the tube was spun down and the aqueous solvent was decanted. Residual water was further removed slowly in a dehydration chamber. After 48 hours, the dehydrated fibers were cryo-protected in 30% glycerol and flash-frozen in liquid nitrogen. The fiber X-ray diffraction pattern was collected using an in-house sealed tube microfocus X-ray generator (MicroMax-003, Rigaku Corporation) with Saturn944 CCD and Cu-Kα (1.54 Å) radiation at 100 K. This result is shown in Figure 3B. In another experiment, TDP-43(341-357) fibers were prepared as described above, except that they were left in the dehydration chamber for 24 hours. The resulting fibers were analyzed with an X-ray

powder diffractometer (PANalytical X'Pert 3) on a spinning sample holder. The data were collected over a 2θ range from 1º to 50º. This result is shown in Figure S10.

4.5.2 Computational methods

MD and REMD Simulations. All simulations were performed using amber99sb-ildn force field parameters[42] as implemented in the GROMACS package[43], version 4.5.5. The systems were placed in TIP3P water[44] cubic boxes and simulated under NpT conditions. All the individual runs ranged from 100 to 110 ns. The peptide model included N-terminal acetyl and C-terminal amide groups to avoid artificially favoring β-hairpin formation through attractive electrostatic interactions. For specific details we refer the reader to the SI Methods.

Additionally, implicit-solvent Replica-Exchange Molecular Dynamics (REMD)[45] simulations were carried out to enhance sampling. Temperatures were exponentially separated ranging from 280 to 450K, using 6 replicas (50 ns each replica for a total of 300 ns), and with exchange probabilities ranging from 0.3 to 0.5. The REMD simulation was initiated using 2 TDP-43(341-357) in the β-hairpin conformation placed at random starting positions. Molecular Dynamics (with 50 ns per replica) from NVT pre-equilibrated systems were performed, and all the interactions were calculated by setting a cut-off distance of 0, which for GROMACS is equivalent to computing all interactions using infinite cut-off. Additional computational details and a figure (Fig. S12) illustrating the starting peptide configuration are given in the SI.

Protein-Protein Docking. Docking of two ten-stranded systems was performed with the HADDOCK[32] program, using data from solvent accessible surface (SAS) calculations. The HADDOCK program was used to generate the twenty-stranded oligomer with side chain intermeshing (dry interface), made up by a pair of pentamer TDP-43 (each sheet contains five β-hairpins and thus ten β-strands). For specific details we refer the reader to the SI Methods.

Semiempirical QM Calculations. To study the energetics of oligomer formation, AM1[46] calculations were performed in gas-phase using the Gaussian09[47] suite package (Revision D.01). The choice of a semi-empirical method was based on the differences in size from the smallest to the biggest system, which span up to ten-stranded β-sheets (corresponding to five TDP-43(341-357) β-hairpins). Figure S9 and Table S4 contains a complete description of the procedures used.

4.6 Acknowledgments

We thank Dr. Margarita Menendez for use of the Cary 210 Spectrophotometer This work was supported by Grants CTQ2010-21567-C02-02 (FPI Fellowship to MMG), SAF2013-49179-C2-2-R (DVL) and SAF2013-49179-C2-1-R (MCV) and an EU JPND AC14/00037 (DVL and MCV), AriSLA TARMA (FB), Thierry Latran Foundation REHNPALS (EB), the EU Joint Programme-Neurodegenerative Diseases JPND RiMod-FTD, Italy, Ministero della Sanita' (EB).

MMG is thankful for the kind hospitality of the members of the McDermott and Tong laboratories during his stay in NY during the summer of 2014 (funded by FPI fellowship CTQ2010-21567-C02-02).

4.7 References

(1) Buratti, E.; Baralle, F. E. TDP-43: Gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem. Sci.* **2012,** *37,* 237-247.

(2) Neumann, M.; Sampathu, D. M.; Kwong, L. K.; Truax, A. C.; Micsenyi, M. C.; Chou, T. T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C. M.; McCluskey, L. F.; Miller, B. L.; Masliah, E.; Mackenzie, I. R.; Feldman, H.; Feiden, W.; Kretzschmar, H. A.; Trojanowski, J. Q.; Lee, V. M. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006,** *314,* 130-133.

(3) Josephs, K. A.; Whitwell, J. L.; Weigand, S. D.; Murray, M. E.; Tosakulwong, N.; Liesinger, A. M.; Petrucelli, L.; Senjem, M. L.; Knopman, D. S.; Boeve, B. F.; Ivnik, R. J.; Smith, G. E.; Jack, C. R. Jr.; Parisi, J. E.; Petersen, R. C.; Dickson, D. W. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol.* **2014,** *127,* 811-824.

(4) Zhu, L.; Xu, M.; Yang, M.; Yang, Y.; Li, Y.; Deng, J.; Ruan, L.; Liu, J.; Du, S.; Liu, X.; Feng, W.; Fushimi, K.; Bigio, E. H.; Mesulam, M.; Wang, C.; Wu, J. Y. An ALS-mutant TDP-43 neurotoxic peptide adopts an anti-parallel β-structure and induces TDP-43 redistribution. *Hum. Mol. Genet.* **2014**, *23,* 6863-6877.

(5) Wang, X.; Blanchard, J.; Grundke-Igbal, I.; Wegiel, J.; Deng, H. X.; Siddique, T.; Igbal, K. Alzheimer disease and amyotrophic lateral sclerosis: an etiopathogenic connection. *Acta Neuropathol.* **2014,** *127,* 243-256.

(6) Arai, T.; Hasegawa, M.; Akiyama, H.; Ikeda, K.; Nonaka, T.; Mori, H.; Mann, D.; Tsuchiya, K.; Yoshida, M.; Hashizume, Y.; Oda, T. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem. Biophys. Res. Comm.* **2006,** *351,* 602-611.

(7) Robinson, J. L.; Geser, F.; Stieber, A.; Umoh, M.; Kwong, L. K.; Van Deerlin, V. M.; Lee, V. M.; Trojanowski, J. Q. TDP-43 skeins show properties of amyloid in a subset of ALS cases. *Acta Neuropathol.* **2013,** *125,* 121-131.

(8) Bigio, E. H.; Wu, J. Y.; Deng, H. X.; Bit-Ivan, E. N.; Mao, Q.; Ganti, R.; Peterson, M.; Siddique, N.; Geula, C.; Siddique, T.; Mesulam, M. Inclusions in frontotemporal lobar

degeneration with TDP-43 proteinopathy (FTLD-TDP) and amyotrophic lateral sclerosis (ALS), but not FTLD with FUS proteinopathy (FTLD-FUS), have properties of amyloid. *Acta Neuropathol*. **2013,** *125,* 463-465.

(9) Chen, A. K. H.; Lin, R. Y. Y.; Hsieh, E. Z. J.; Tu, P. H.; Chen, R. P.; Liao, T. Y.; Chen, W.; Wang, C. H.; Huang, J. J. Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc*. **2010,** *132,* 1186-1187.

(10) Saini, A.; Chauhan, V. S. Delineation of the core aggregation sequences of TDP-43 C-terminal fragment. *Chembiochem*. **2011,** *12,* 2495-2501.

(11) Guo, W.; Chen, Y.; Zhou, X.; Kar, A.; Ray, P.; Chen, X.; Rao, E. J.; Yang, M.; Ye, H.; Zhu, L.; Liu, J.; Xu, M.; Yang, Y.; Wang, C.; Zhang, D.; Bigio, E. H.; Mesulam, M.; Shen, Y.; Xu, Q.; Fushimi, K; Wu, J. Y. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol*. **2011,** *18,* 822-830.

(12) Sun, C. S.; Wang, C. Y.; Chen, B. P.; He, R. Y.; Liu, G. C.; Wang, C. H.; Chen, W.; Chern, Y.; Huang, J. J. The influence of pathological mutations and proline substitutions in TDP-43 glycine-rich peptides on its amyloid properties and celular toxicity. *PLoS One* **2014,** 9(8):e103644.

(13) Saini, A.; Chauhan, V. S. Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir* **2014,** *30,* 3845-3856.

(14) Fang, Y. S.; Tsai, K. J.; Chang, Y. J.; Kao, P.; Woods, R.; Kuo, P. H.; Wu, C. C.; Liao, J. Y.; Chou, S. C.; Lin, V.; Jin, L. W.; Yuan, H. S.; Cheng, I. H.; Tu, P. H.; Chen, Y. R. Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun*. **2014**, *5,* 4824.

(15) Budini, M.; Romano, V.; Quadri, Z.; Buratti, E.; Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum. Mol. Genet*. **2015,** *24,* 9-20.

(16) Teng, P. K.; Anderson, N. J.; Goldschmidt, L.; Sawaya, M. R.; Sambashivan, S.; Eisenberg, D. Ribonuclease A suggests how proteins self-chaperone against amyloid fiber formation. *Protein Sci*. **2012**, *21,* 26-37.

(17) Budini, M.; Buratti, E.; Stuani, C.; Guarnaccia, C.; Romano, V.; De Conti, L.; Baralle, F. E. Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich región. *J. Biol. Chem*. **2012,** *287,* 7512-7525.

(18) Mompeán, M.; Buratti, E.; Guarnaccia, C.; Brito, R. M. M.; Chakrabartty, A.; Baralle, F. E.; Laurents, D. V. Structural characterization of the minimal segment of TDP-43 competent for aggregation. *Arch. Biochem. Biophys*. **2014,** *545,* 53-62.

(19) Kar, K.; Hoop, C. L.; Drombosky, K. W.; Baker, M. A.; Kodali, R.; Arduini, I.; van der Wel, P. C.; Horne, W. S.; Wertzel, R. β-hairpin mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol*. **2013,** *425,*1183-97.

(20) Kirschner, D. A.; Inouye, H.; Duffy, L. K.; Sinclair, A.; Lind, M.; Selkoe, D. J. Synthetic peptide homologous to beta protein from Alzheimer disease forms amyloid-like fibrils *in vitro*. *Proc. Natl. Acad. Sci. USA* **1987,** *84,* 6953-6957.

(21) Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **2003,** *300,* 486-489.

(22) Kar, K.; Jayaraman, M.; Sahoo, B.; Kodali, R.; Wetzel, R. Critical nucleus size for disease-related polyglutamine aggregation is repeat length dependent. *Nat. Struct. Mol.*

Biol. **2011,** *18,* 328-336.
(23) Lashuel, H. A.; Hartley, D. M.; Petre, B. M.; Wall, J. S.; Simon, M. N.; Walz, T.; Lansbury, P. T. Jr. Mixtures of wild-type and a pathogenic (E22G) form of Abeta40 in vitro accumulate proofibrils, including amyloid pores. *J. Mol. Biol*. **2003,** *332,* 795-805.
(24) Wang, Y. T.; Kuo, P. H.; Chiang, C. H.; Liang, J. R.; Chen, Y. R.; Wang, S.; Shen, J. C.; Yuan, H. S. The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem*. **2013,** *288,* 9049-9057.
(25) Kayed, R.; Head, E.; Sarsoza, F.; Saing, T.; Cotman, C. W.; Necula, M.; Margol, L.; Wu, J.; Breydo, L.; Thompson, J. L.; Rasool, S.; Gurlo, T.; Butler, P.; Glabe, C. G. Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Mol. Neurodegener*. **2007,** *2,* 18.
(26) Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*. **1993,** *26,* 283-291.
(27) MacPhee, C. E.; Dobson, C. M. Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *J. Mol. Biol*. **2000,** *297,* 1203-1215.
(28) Hervás, R.; Oroz, J.; Galera-Prat, A.; Goñi, O.; Valbuena, A.; Vera, A. M.; Gómez-Sicilia, A.; Losada-Urzáiz, F.; Uversky, V. N.; Menéndez, M.; Laurents, D. V.; Bruix, M; Carrión-Vázquez, M. Common features at the start of the neurodegeneration cascade. *PLoS Biol*. **2012,** 10(5):e1001335.
(29) Tsemekhman, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci*. **2007,** *16,* 761-764.
(30) Mompeán, M.; González, C.; Lomba, E.; Laurents, D. V. Combining Classical MD and QM calculations to elucidate complex system nucleation: a twisted, three-stranded, parallel β-sheet seeds amyloid fibril conception. *J. Phys. Chem. B* **2014,** *118,* 7312-7316.
(31) Rosseti, G.; Magistrato, A.; Pastore, A.; Carloni, P. Hydrogen bonding cooperativity in polyQ β-sheets from first principle calculations. *J. Chem. Theory Comput*. **2010,** *6,* 1777-1782
(32) Dominguez, C; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc*. **2003,** *125,* 1731-1737.
(33) Lawrence, M. C.; Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol*. **1993**, *234*, 946-950.
(34) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **2005,** *435,* 773-778.
(35) Sharma, D.; Shinchuk, L. M.; Inouye, H.; Wetzel, R.; Kirschner, D. A. Polyglutamine homopolymers having 8-45 residues form slablike beta-crystallite assemblies. *Proteins* **2005,** *61,* 398-411.
(36) Buchanan, L. E.; Carr, J. K.; Fluitt, A. M.; Hoganson, A. J.; Moran, S. D.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc. Natl. Acad. USA* **2014,** *111,* 5796-5801.

(37) Anthis, N. J.; Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci*. **2013,** *22,* 851-858.
(38) Nagai, Y.; Tucker, T.; Ren, H.; Kenan, D. J.; Henderson, B. S.; Keene, J. D.; Strittmatter, W. J.; Burke J. R. Inhibition of Polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J. Biol. Chem*. **2000**, *275,* 10437-10442.
(39) LeVine, H. 3rd. Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci*. **1993,** *2,* 404-410.
(40) Klunk, W. E.; Pettegrew, J. W.; Abraham, D. J. Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J. Histochem. Cytochem*. **1989,** *37,* 1273–1281.
(41) Diez-García, F.; Pantoja-Uceda, D.; Jiménez, M. Á.; Chakrabartty, A.; Laurents, D. V. Structure of a simplified β-hairpin and its ATP complex. *Arch. Biochem. Biophys*. **2013,** *537,* 62-71.
(42) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Kepleis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010,** *78,* 1950-1958.
(43) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput*. **2008,** 4, 435-447.
(44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. **1983**, *79,* 926-935.
(45) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett*. **1999,** *314,* 141-151.
(46) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985,** *107,* 3902-3909.
(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; *et al*. Gaussian 09, revision D.01; Gaussian, Inc.: Wallingford, CT, 2009.

4.8 Supporting Information

4.8.1 Supporting Methods

Circular Dichroism: Secondary Structure Analysis.

Estimations of secondary structure content were performed using the CDSSTR algorithm[1], as implemented on the webserver Dichroweb*:* ***http://www.cryst.bbk.ac.uk/cdweb*** [2,3].

NMR Spectroscopy analysis of the longer peptide TDP-43 (322-366)

The peptide TDP-43(322-366) was synthesized by standard solid phase methodology and purified at the ICGEB, Trieste. It was dissolved in aqueous buffer containing 100% D2O for the 2D 1H-13C HSQC spectrum or alternatively 90% H2O/10% D2O for the 1D 1H, 1H 2D 1H TOCSY and NOESY spectra. Both samples contained 1 mM deuterated acetic acid and the pH was 6.0. All spectra were recorded 5ºC on a Bruker 600 MHz NMR spectrometer equipped with a cryo probe and z-gradients. For the TOCSY and NOESY spectra, mixing times of 60 ms and 150 ms were used, respectively, and the matrix was 2k x 512. The spectra were acquired and processed using the Bruker program Topspin version 2.1.

Molecular Dynamics (MD) and Replica-Exchange MD (REMD) Simulations

All the simulations were performed using the amber99sb-ildn force field parameters[4] as implemented in the GROMACS package[5], version 4.5.5. The choice of this force field was based on previous satisfactory results describing amyloid-like systems[6,7]. The systems were placed in TIP3P water[8] cubic boxes, leaving at least 1.2 nm between each solute atom and box edges. After steepest descent energy minimization, two equilibration periods were performed restraining the protein atoms movement, to allow solvent molecules equilibrate around the solute (500ps, NVT and 2ns NpT). The Berendsen[9] weak coupling scheme was used during these periods. For the MD data collection stage, we used the Nosé-Hoover[10] thermostat and Parrinello-Rahman[11] barostat in absence of any restraints, with time constants of 0.5 and 1.0 ps, respectively. The LINCS algorithm[12] was applied in all the simulations to allow a time step of 2 fs. All the individual runs ranged from 100 to 110 ns. The peptide model included N-terminal acetyl and C-terminal amide groups to avoid artificially favoring β-hairpin formation through attractive electrostatic interactions. To ensure the independence of the simulations, we used different initial velocities in each run. The short-range nonbonded interactions were cut-off at 1 nm, while long-range electrostatics were calculated with the particle mesh Ewald (PME) algorithm[13]. Dispersion-correction was applied to account for van der Waals interactions at distances longer than the cut off, and periodic boundary conditions were applied in all directions. The peptide models were built using PyMOL (http://www.pymol.org/).

Protein-Protein Docking

Docking of two ten-stranded systems was performed with the HADDOCK[14] program, using data from solvent accessible surface (SAS) calculations. The HADDOCK program was used to generate the twenty-stranded oligomer with sidechain intermeshing (dry interface), made up by a pair of pentamer TDP-43 (each sheet contains five β-hairpins and thus ten β-strands). The residues with high solvent accessibility (>45%) were selected as active residues and the neighbors of these active residues were selected as passive residues. After rigid body-energy minimization, the best 200 solutions based on the inter-molecular energy were selected for the semiflexible docking, followed by explicit water solvation. The top 200 docked models were clustered by a cutoff of 3.5 Å, with a minimum of 10 structures in each cluster, which yielded six clusters. In terms of HADDOCK score and total energy, the best 10 structures from the first cluster were selected as the final model for the TDP-43 (341-357) oligomer.

4.8.2 Supporting Figures

The NMR spectra of the TDP-43 (322-366) show poor chemical shift despersion, which is typical of random coil ensembles. In a previous study of the shorter peptide TDP-43 (342-366), the NMR signals were assigned unambiguously and this segment was determined to exist (prior to aggregation) as a random coil (Mompean et al. (2014) ABB). Here, we find no significant chemical shift differences for the signals arising from residues 342-366 in the shorter peptide compared to the longer peptide spanning residues 322-366. This is strong evidence that this segment still exists initially as a random coil ensemble in the context of the longer 322-366 peptide. The longer peptide contains many Ala (7) and Met (5) residues which we could not assign individually due to strong overlap. This lack of dispersion and the chemical shift values of the overlapped Ala and Met resonances are consistent with a random coil ensemble containing a minor (<25%) population of helical confomers.

The sole Trp residue at position 334 was completely assigned. Its chemical shift values (in ppm): H=8.17, Ha=4.63, HB2=3.32, HB3=3.27, HD=7.20, HNE=10.21, HE3=7.48, HZ2=7.48, HZ3=7.10, HH2=7.25, CA=57.4, CB=29.4, are all close to those observed for Trp residues in short, unstructured peptides. Furthermore, we observe no NOE signals between the aromatic Trp H and Gln 327 or with aliphatic H from other residues in this polypeptide.

1D 1H NMR spectra recorded confirm that the TDP-43 (322-366) concentration was unchanged and that no aggregation occurred during the acquisition of these 2D spectra (18 h); conditions in which TDP-43 (341-366) and TDP-43(341-357) do extensively aggregate and lose most of their NMR signal intensity. The TDP-43 (322-366) sample did partially precipitate after 40 h. Taking these findings into account, we conclude that flanking segments do not promote the aggregation of TDP-43 (341-357).

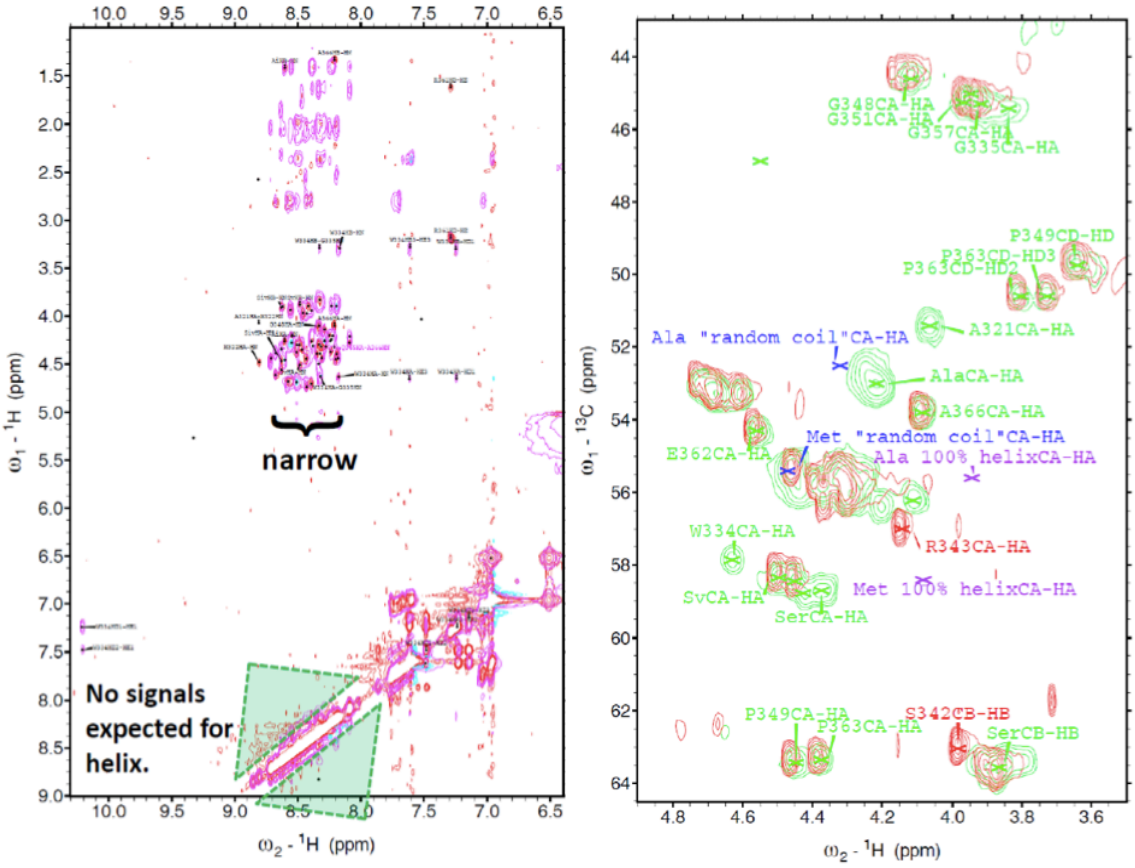


Figure S1: NOESY (left panel) and ^{1}H-^{13}C HSQC (right panel) NMR spectra of TDP-43 (322-366). Downfield region of the 2D 1H NOESY NMR spectrum of TDP-43 (322-366) (***left***). The narrow chemical shift dispersion of the ^{1}HN signals (<1 ppm) is evident. The region where $^{1}HN_{i}$ to $^{1}HN_{i+1}$, $^{1}HN_{i+3}$ and $^{1}HN_{1+4}$ NOE signals characteristically appear in helical peptides and proteins is indicated. No such signals are observed here. The alpha region of the ^{1}H-^{13}C HSQC spectra of TDP-43(322-366) (in **green**) and TDP-43 (342-366) Q343R variant (in **red**) (***right***). Some nuclei are labeled. "X's" mark the expected chemical shifts values of Ala and Met HC alpha signals in random coil (**blue**) or 100% helical (**purple**) conformations. It is clear that the group of overlapped Ala peaks is much closer to the position expected for random coil than to that expected for helix.

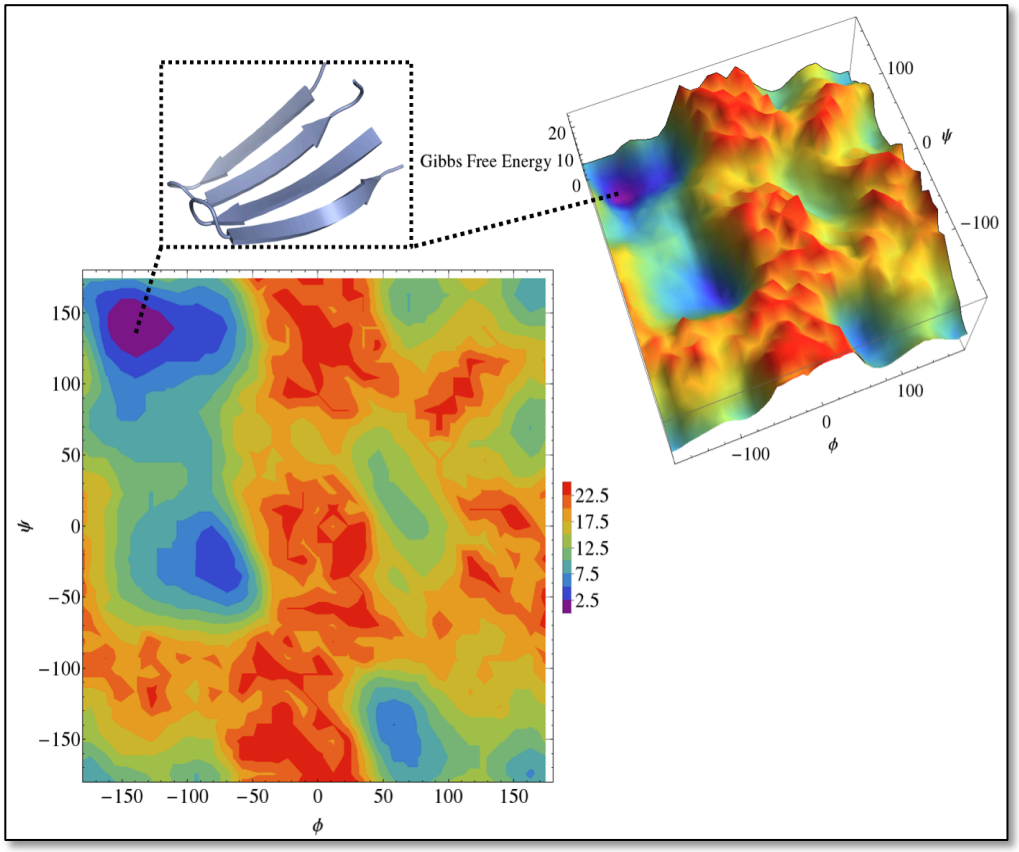


Figure S2: Free Energy Landscapes in the phi and psi dihedral angle space. The phi and psi torsion angles were extracted from the trajectory of a REMD simulation of two TDP-43 dimers with the starting configuration used for the MD runs. The free energy landscape in this set of angles was built from the g_sham utility implemented in GROMACS. The time of the trajectory at which this minimum occurs can be obtained directly from the coordinates (phi and psi) of this point, which is equivalent to an ***in-register,*** lateral association. Since these angles define the secondary structure of the peptides, the 2D projection of the landscape corresponds to the Ramachandran plot, where it can be seen that the minimum lies in the β-sheet region.

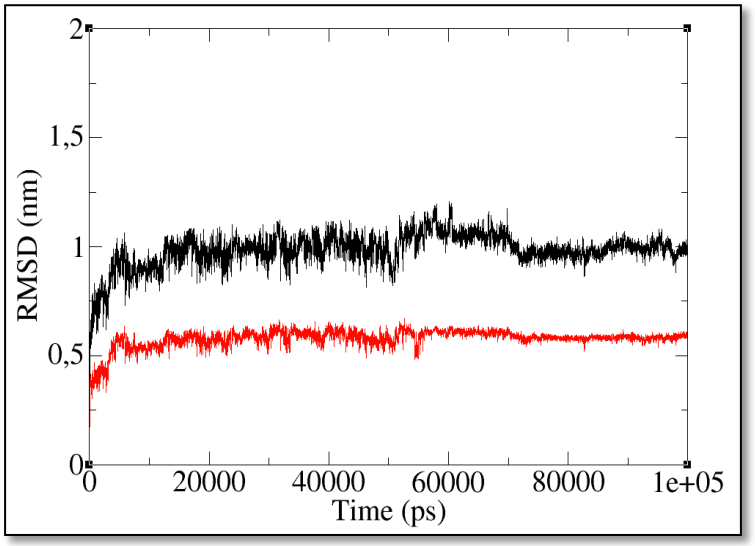


Figure S3: dRMSD for the lateral association (I). The **black** line corresponds to that of the same color in **Fig. 1** of the main text (*out-of-register* binding mode), representing the standard RMSD. The dRMSD is shown in red, which displays a lower value but the same trend as expected. As in Fig. 1, the dRMSD is calculated with respect to the time-zero structure.

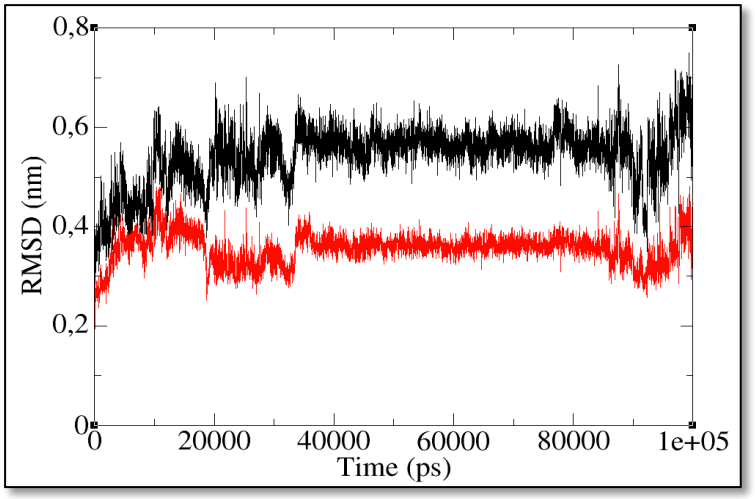


Figure S4: dRMSD for the lateral association (II). The **black** line corresponds to the RMSD of the first *in-register* binding mode simulation (red curve of **Fig. 1** in the main text). As in **Fig. S2**, the red color is that of the dRMSD and the calculations are performed with respect to the time-zero structures.

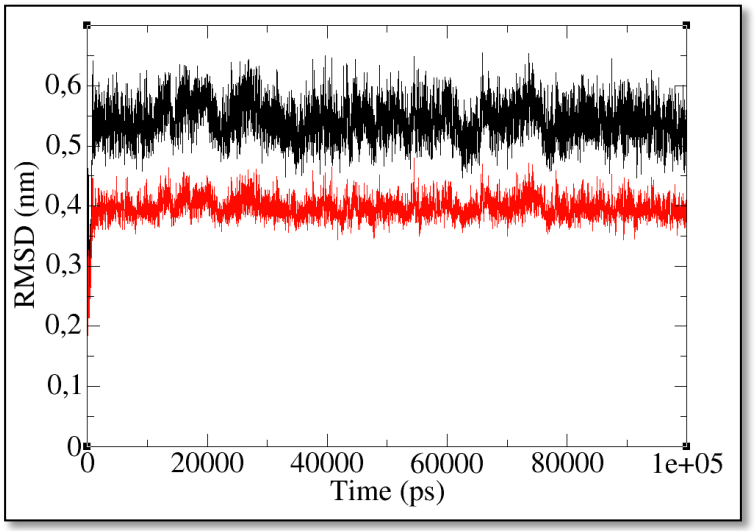


Figure S5: dRMSD for the lateral association (III). The **black** line corresponds to the RMSD of the second in-register binding mode simulation (green curve of **Fig. 1** in the main text). The related dRMSD analysis is shown here in red, as in **Fig. S2** and **Fig. S3**. Again, all the calculations were performed with respect to the time-zero structures.

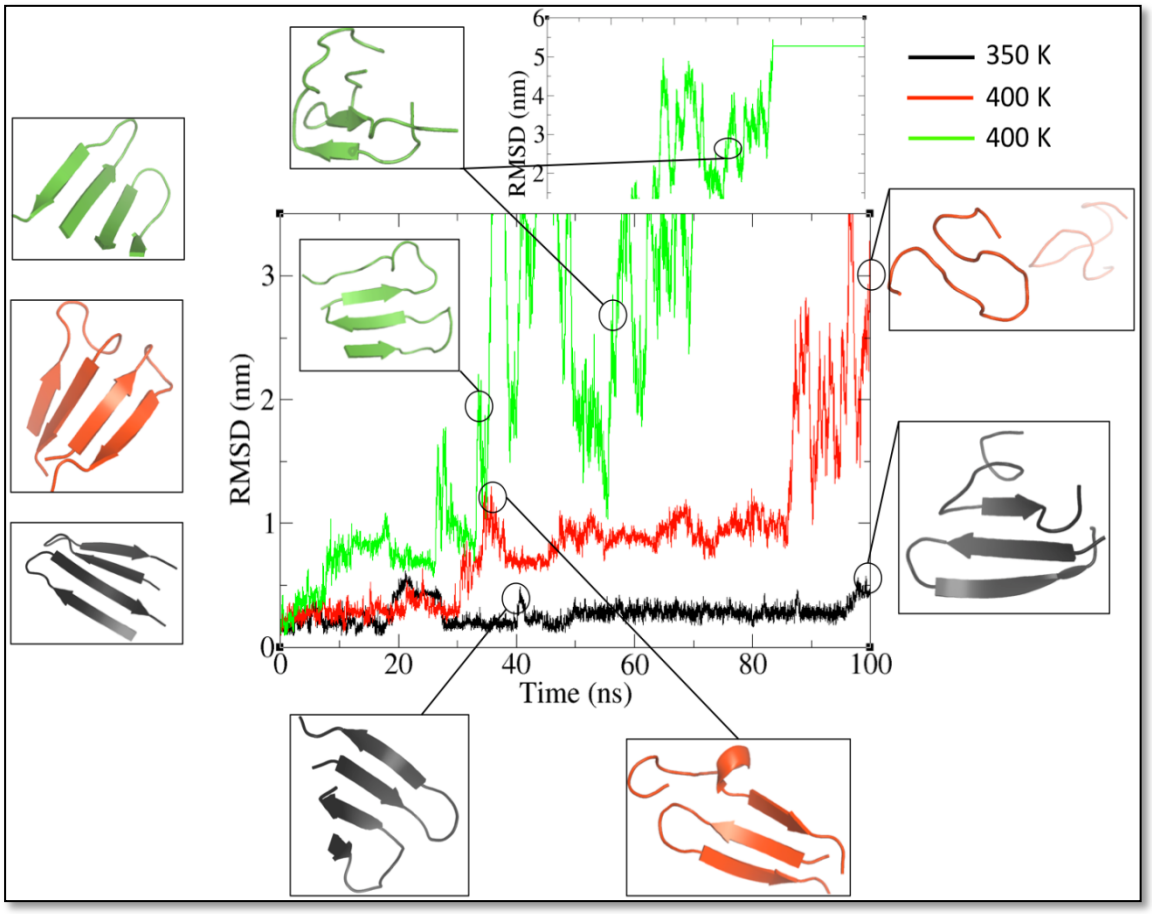
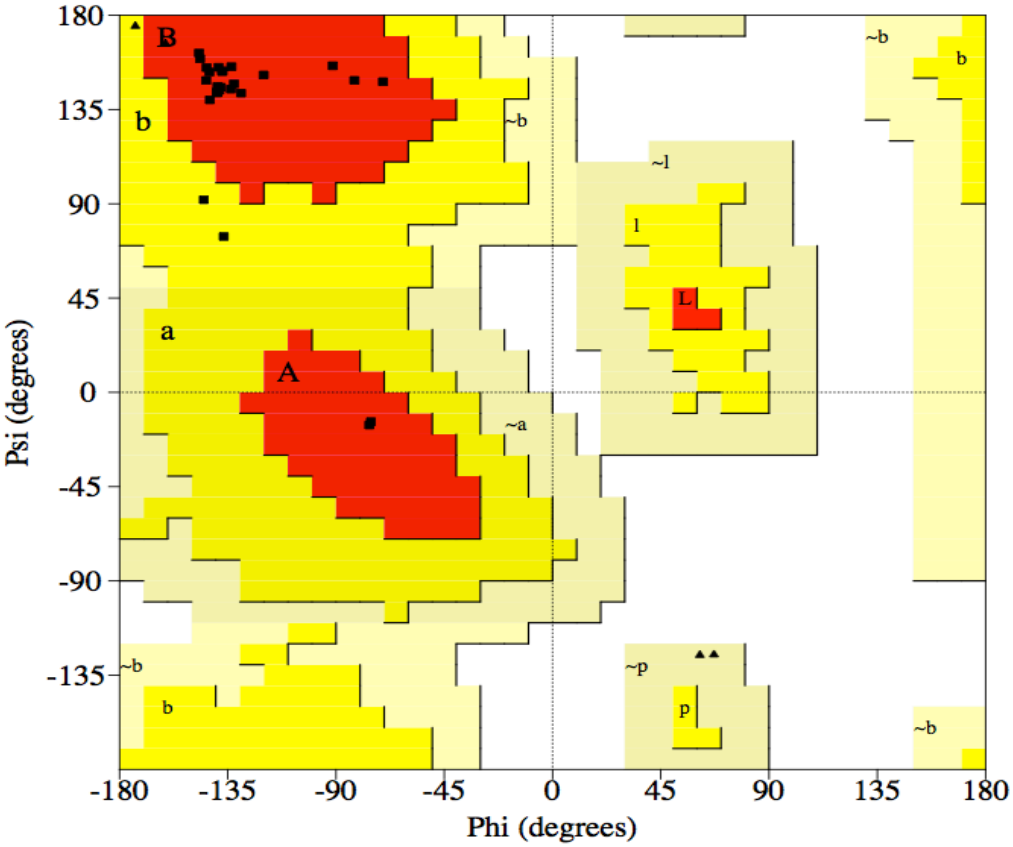


Figure S6: *In-Silico* thermal stability. Trajectories from simulations at high temperature revealing unfolding preferences. The starting structures are depicted on the far left, whereas intermediate and final snapshots are illustrated with circles corresponding to the frames where the snapshots were taken. The RMSD profiles were calculated with respect to the corresponding time-zero structures.



Plot statistics

Residues in most favoured regions [A,B,L]	22	91.7%
Residues in additional allowed regions [a,b,l,p]	2	8.3%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
	----	------
Number of non-glycine and non-proline residues	24	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	2	

Total number of residues	34	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure S7: PROCHECK structure validation. A (representative/average) TDP-43(341-357) dimer found in the MD simulations was subjected to stereochemical structure validation. The statistical analysis revealed that 91.7% of the residues are in the most favored region and the remaining 8.3% are also plausible. None of the residues are found in regions generously allowed or disallowed regions. The PROCHECK criterion for a structure to be valid is that at least 90% of the residues have to be present in the most favored regions, which is fulfilled for this system.

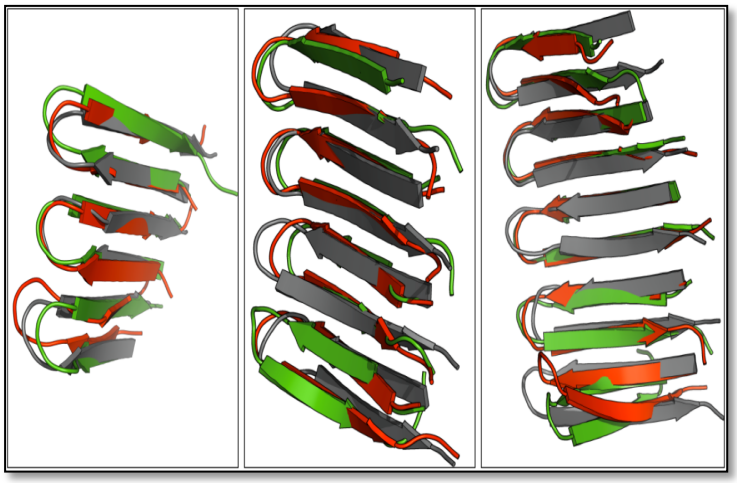


Figure S8: Structural stability of the TDP-43(341-357) oligomers. Structural stability of a TDP-43 (341-357) trimer (*left panel*), tetramer (*central panel*) and pentamer (*right panel).* Snapshots from the oligomers at time 0 (grey), 50 (red) and 90 ns (green) from each trajectory. Alignment and superposition of these structures highlight their stability over the course of the 100 ns simulations.

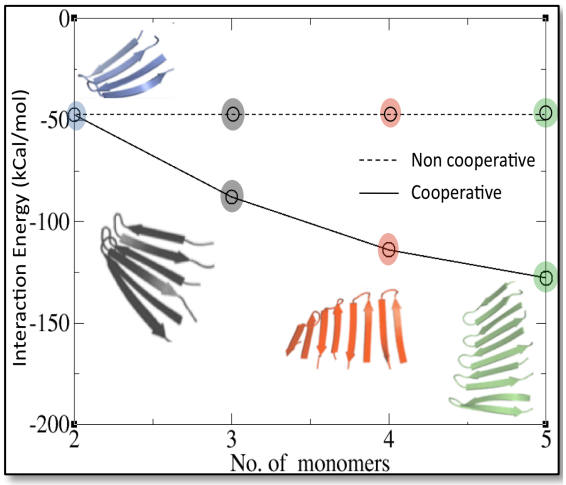


Figure S9: Cooperative hydrogen bonding formation. The interaction energy for two to five TDP-43(341-357) β-hairpins is shown. If the lateral association process associated to fibril formation were not cooperative, then the energy released by oligomerization would be less than the linearly expected, and this is shown in the black line. Each system (dimer, trimer, tetramer and pentamer) is represented in a color code (blue, **black**, red, and green, respectively). For the sake of clarity, their corresponding energy values are indicated with the same color within the plot. These values are shown in **Sup. Table 3**.

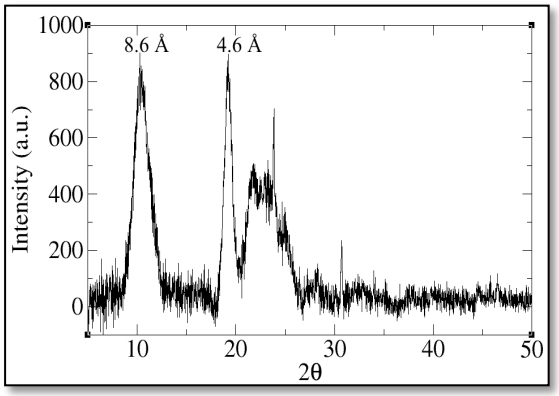


Figure S10: TDP-43(341-357) X-Ray fiber diffraction (XRFD). The X-ray diffraction pattern of preformed fibrils shows the same distances that were observed in the crystal pattern, corresponding to the space between two β-strands in the same β-sheet and between two facing β-strands in different β-sheets facing each other at the dry interface.

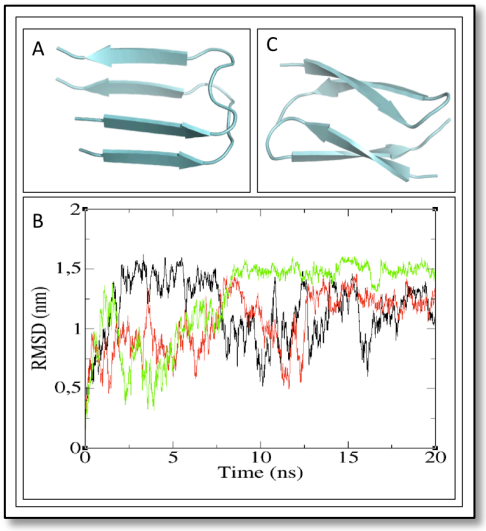


Figure S11: β-arc and β-turn topologies. (A) In the "β-arc" topology each TDP-43(341-357) β-hairpin contributes one β-strand to two different β-sheets. Formation of a dry interface restrained to this topology involves longer loops and as a result, significant structural distortion occurs and RMSD values as high as 1.5 nm are reached after only 20 ns, as represented in panel **(B)** Each line (black, green, and red) corresponds to an independent simulation with the same starting configuration (represented in panel (C)) but different initial velocities. All the RMSD curves were calculated with respect to the corresponding time-zero structures. **(C)** In contrast, β-hairpin formation is facilitated by the presence of a Gly-Pro turn. In this "β-turn" topology each β-hairpin associates in parallel to contribute two β-strands to the same β-sheet. The structural stability of this motif as the kernel of TDP-43(341-357) aggregation is addressed throughout the main text.

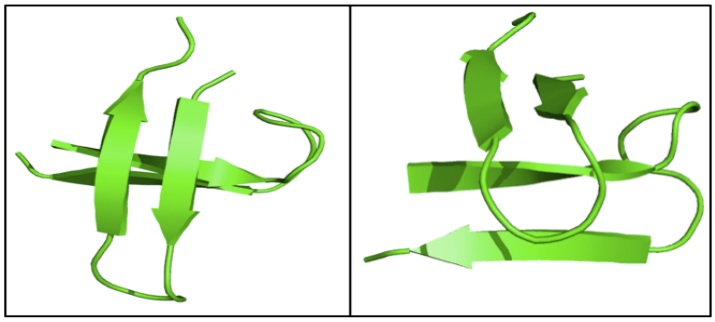


Figure S12. Initial Configuration used in the Simulations. Two different views of two beta-hairpins stacked with an antiparallel orientation with respect to each other at ~ 4.5 - 5 Å, as we observed this inter-strand backbone distances by X-Ray. This set up allows for β-sheet formation within 100 ns MD simulations.

4.8.3 Supporting Tables.

Supplementary Table S1: Peptides Derived from TDP-43 and Reported to Form Aggregates.

Study's Reference	Residue # in TDP-43	Techniques Utilized	Main Results
Guo, W....Wu JY *et al* (2011) Nat. Struct. Mol. Biol. 18:822-831	286-331 & variant with A315T	ThT, AFM, EM, gel filtration, cytotoxicity assays,	Forms an amyloid-like fibril that is neurotoxic. The A315T mutation enhances fibril formation and toxicity.
Zhu L....Wu, JY *et al*. (2014) Hum. Mol. Genetics 23:6863-6877	286-331 & variant with A315T	EM, AMF, ThT, CD, NMR, in cell fluorescence	M307-N319 would adopt a β-strand that packs in an antiparallel fashion. The A315T variant causes the protein's in cell localization to change. This variant can promote the aggretation of Abeta1-40
Liu GCH ...Huang JJT (2013) Chem. Comm. 49:11212-4	287-322, 292-322, 297-322, 302-322, 307-322, 307-322sc, 312-322	EM, CD, vesicle disruption assay	Peptides 287-322, 292-322, 297-322, 302-322, 307-322, but not 312-322 form amyloid-like fibrils. 307-322, but not scrambled 307-322 can induce the aggregation of full length TDP-43
Sun CS … Huang JJT *et al*. (2014) PloS One 9: e103644	287-322 and variants: G294A, G294V, G295S, G294P, G294,5,6P, G208,9,10P	EM, CD, ThT, FT-Raman, cytotoxicity assays	287-322 and its variants G294A, G294V, G295S form amyloid-like fibrils. The formation of fibrils is reduced or completely disrupted in the other variants.
Jiang *et al.* (2013) J. Biol. Chem. 288:19614-19624	318-343	ThT fluorescence, CD, NMR, electrophoresis,	Segments 321-330 and 335-343 said to form transient helices which would later convert into beta-hairpins that pack in an anti-parallel fashion.
Wang YT... Yuan HS *et al* (2013) J. Biol. Chem. 288:9049-9057	208-414	SAXS, EM, CD, ThT	Truncated the RRM2 domain + C-terminal region. Propose that 2 beta strands from this cut RRM2 domain could form amyloid, thereby disrupting the native dimeric structure of TDP-43, which is due to a head/head interaction of the N-terminal domains.
Chen et al. (2010) JACS 132:1186-1187	287-322, 321-353, 351-383, 381-414	EM, CD, ThT, Cytotoxicity Assay	287-322 forms amyloid-like fibrils as do its variants A315T & G294A. The other peptides form amphorous aggregates. Note that the segment studied by us (341-357) would be divided between 2 of the peptides studied by Chen et al. and therefore it is not surprising that this region does not aggregate.
Saini A & Chauhan VS (2011) ChemBioChem 12:2495-2501	15 12-mer peptides spanning TDP-43 residues 220-414. Then 13 peptides derived from TDP-43 segments 246-258 and 311-328	EM, ThT	Found that 246-255 and 311-320 can form amyloid-like fibrils that enhance ThT fluorescence. In their peptides, the key 341-357 would have been split into two different peptides (337-349/ 340-362)
Saini A & Chauhan VS Langmuir (2014) 30:3845-3856	Small peptides derived from TDP-43 346-355 and 311-320	EM, CD, Theology, light scattering	Found that 246-255 and its smaller fragment DLII, and 311-320 and its smaller fragment NFGAF can form amyloid-like fibrils and hydrogels.

Supplementary Table S2: Secondary Structure from Analysis of Circular Dichroism Spectra of TDP-43 341-357 peptides at 25 ºC, pH 6.8 utilizing the CDSSTR algorithm.

	% Secondary Structure			
	5 ºC		25 ºC	
	Ref. Set 4	Ref. Set 7	Ref. Set 4	Ref. Set 7
α-helix	18	31	16	23
β-sheet	26	19	28	23
turn	25	19	25	19
random coil	31	32	32	36
RMSD	0.03	0.04	0.03	0.03

Supplementary Table S3: Interaction energy values, related to Fig. S9. The coordinates of the different systems were obtained from the time-averaged structure of largest oligomer (*i.e.*, the pentamer) over a 100 ns MD simulation (See Fig. 2B of the main text). The energy values reported correspond to systems that were relaxed by restraining the Cα positions to their experimental interstrand distance values of 4.6 Å. The interaction energies for each system were obtained from the AM1 energy values of the dimer, trimer, tetramer, and pentamer, with respect to the energy of a monomer times the number of monomers in each oligomer.

No. of monomers	**ΔE_{int} (kcal/mol)**
2	-47.34
3	-87.80
4	-113.86
5	-127.87

Supplementary Table S4: X-Ray diffraction reflections in known amyloid and amyloid-like fibrils.

X-ray fibril diffraction reflections

Polypeptide	axial reflection Å (intra-β-sheet)	equatorial reflection Å (inter-β-sheet)	Ref.
TDP-43 (341-357)	4.6	8.6	This work
polyQ	4.7	8.2	(15)
Abeta	4.7	10	(16)
Abeta	4.7	10.4	(17)
α-syn	4.7	10 – 11	(18)
Adenovirus fibril	4.8	11.1	(19)
β2-microglobulin	4.8	11.7	(20)
CPEB (*Aplysia*)	4.8	10.7	(21)
Insulin	4.8	9.6	(22)
PI3-SH3	4.7	9.4	(23)
Sup35	4.7	10	(24)

4.8.4 Supporting References

(1) Johnson WC (1999) Analysing protein circular dichroism spectra for accurate secondary structures. *Proteins* 35, 307-312.

(2) Whitmore L, Wallace BA (2008) Protein secondary structure analysis from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89, 392-400.

(3) Whitmore L., Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* 32:W668-673.

(4) Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Kepleis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78, 1950.

(5) Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4, 435-447.

(6) Berhanu WM, Hansmann UHE (2012) Side-chain hydrophobicity and the stability of Abeta 16-22 aggregates. *Protein Science* 21, 1837-1848.

(7) Mompeán M, González C, Lomba E, Laurents DV (2014) Combining Classical MD and QM calculations to elucidate complex system nucleation: a twisted, three-stranded, parallel β-sheet sedes amyloid fibril conception. *J Phys Chem B* 118, 7312-6.

(8) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79, 926-935.

(9) Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684.

(10) Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52, 255-268.

(11) Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52, 7182.

(12) Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18, 1463-1472.

(13) Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N·lon(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089-10092.

(14) Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J Am Chem Soc* 125, 1731-1737.

(15) Sharma D, Shinchuk LM, Inouye H, Wetzel R, Kirschner DA (2005) Polyglutamine homopolymers having 8-45 residues form slablike beta-crystallite assemblies. *Proteins* 61, 398-411.

(16) Petkova AT, Buntkowsky G, Dyda F, Leapman RD, Yau WM, Tycko R (2004) Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J Mol Biol* 335, 247-60.

(17) McDonald M, Box H, Bian W, Kendall A, Tycko R, Stubbs G (2012) Fiber diffraction data indicate a hollow core for the Alzheimer's aβ 3-fold symmetric fibril. *J Mol Biol* 423, 454-61.

(18) Serpell LC, Berriman J, Jakes R, Goedert M, Crowther RA (2000) Fiber diffraction of

synthetic alpha-synuclein filaments shows amyloid-like cross-beta conformation. *Proc Natl Acad Sci USA* 97, 4897-902.
(19) Papanikolopoulou K, Schoehn G, Forge V, Forsyth VT, Riekel C, Hernandez JF, Ruigrok RW, Mitraki A (2005) Amyloid fibril formation from sequences of a natural beta-structured fibrous protein, the adenovirus fiber. *J Biol Chem* 280, 2481-90.
(20) Ivanova MI, Sawaya MR, Gingery M, Attinger A, Eisenberg D (2004) An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril. *Proc Natl Acad Sci USA* 101, 10584-9.
(21) Raveendra BL, Siemer AB, Puthanveettil SV, Hendrickson WA, Kandel ER, McDermott AE (2013) Characterization of prion-like conformational changes of the neuronal isoform of Aplysia CPEB. *Nat Struct Mol Biol* 20, 495-501.
(22) Choi JH, May BC, Wille H, Cohen FE (2009) Molecular modeling of the misfolded insulin subunit and amyloid fibril. *Biophys J* 97, 3187-95.
(23) Ghahghaei A, Faridi N (2009) Review: structure of amyloid fibril in diseases. *J BioMed Sci Eng* 2, 345-358.
(24) Balbirnie M, Grothe R, Eisenberg DS (2001) An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc Natl Acad Sci USA* 98, 2375-80.

Complete citation of Gaussian 09

Gaussian 09, Revision **D.01**, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, M. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.

Chapter 5

High Resolution Structural, Stability and Dynamics Elucidation of the N-terminal Domain of TDP-43 Reveals a Stable, Distinct Fold

Reproduced with permission from Romano V., Pantoja-Uceda D., Stuani C., Baralle F E., Buratti E., and Laurents D.V. "High Resolution Structural, Stability and Dynamics Elucidation of the N-terminal Domain of TDP-43 Reveals a Stable, Distinct Fold." Submitted to FEBS Journal.

The past, like the future, is indefinite and exists only as a spectrum of possibilities.
Stephen Hawking

5.1 Abstract

TDP-43 is an RNA transporting and processing protein whose aberrant aggregates are implicated in neurodegenerative diseases. The C-terminal domain of this protein plays a key role in mediating this process. However, the N-terminal domain (residues 1-77) is needed to effectively recruit TDP-43 monomers into this aggregate. Here we report for the first time, the essentially complete ^{1}H, ^{15}N and ^{13}C NMR assignments and the structure of the N-terminal domain determined on the basis of 26 H-bond, 60 torsion angle and 1058 unambiguous NOE structural restraints. The structure consists of an α-helix and six β-strands. Two β-strands form a β-hairpin not seen in the Ubiquitin fold. All Pro residues are in the *trans* conformer and the two Cys are reduced and distantly separated on the surface of the protein. The domain has a well-defined hydrophobic core composed of F35, Y43, W68, Y73 and 17 aliphatic side chains. The fold is topologically similar to the reported structure of Axin 1. The protein is stable and no denatured species are observed at pH 4 and 25 ºC. At 4 kcal/mol, the conformational stability of the domain, measured by H/D exchange is comparable to Ubiquitin (6 kcal/mol). The β-strands, α-helix, and 3 of 4 turns are generally rigid but the loop formed by residues 47-53 is mobile, as determined by model-free analysis of the $^{15}N\{^{1}H\}$NOE, and the translational and transversal relaxation rates.

5.2 Introduction

TDP-43 is a 414-residue protein belonging to the heterogeneous nuclear ribonucleoprotein (hnRNP) family. First discovered for its capacity to bind the TAR element of the AIDS virus RNA [1], TDP-43's true physiological roles in RNA splicing, stability and transport were identified in 2001 [2], and have been recently reviewed [3]. Nine years ago, TDP-43 was found to form aberrant polyubiquitinated, hyperphosphorylated cytosolic aggregates which have been causatively linked to ALS/FTLD [4, 5], and more recently to Alzheimers [6]. The protein contains an N-terminal domain (NTD), two RRM RNA-binding domains and a long, intrinsically disordered C-terminal region. TDP-43, and in particular its N-terminal and RRM domains are well conserved in animals ranging from man to *C. elegans* [7]. The C-terminal region contains three segments enriched with Gly, hydrophobic and Asn/Gln residues, respectively. Considering that almost all the amino acid substitutions arising from mutations linked genetically to pathology are found in this region [8], the C-terminus appears to be a key protagonist in the pathological processes. There is an open debate and contrasting results regarding whether the toxic aggregate is amorphous [4] or amyloid-like [9, 10] and which segment is key for its formation [11-14]. We have recently shown that tandem repeats of the C-terminal segment enriched with Asn and Gln residues can reproduce almost all the disease hallmarks of the full-length protein in cells [15]. In aqueous solution, the peptide corresponding to this segment (341-366) is initially unfolded but then adopts a aggregating conformer that is rich in β-structure [16] with many amyloid-like characteristics [17]. Similar results have been reported independently for a peptide corresponding to residues 331-360 [18].

Notwithstanding the importance of the C-terminal region, the N-terminal domain (NTD), composed of residues 1-77, has recently been shown to be essential for efficient aggregation that depletes the cell of functional TDP-43 [19, 20]. This result prompted us to characterize the structure, conformational stability and dynamics of this domain using NMR spectroscopy. While our work was underway, a low resolution structural model of the backbone structure was published for the NTD construct containing residues 1-80 and a C-terminal His-tag [21]. The model contains five β-strands and an α-helix whose topology matches the ubiquitin fold. The conformational stability of the NTD in the context of Qin *et al.*'s construct is remarkably low, with the majority of the protein being denatured at equilibrium and the authors hypothesize that this low stability is key for the protein's physiological and pathological roles. Despite the significant advance that these results represent, no structure or chemical shift values were deposited and the resolution reported was low. In addition, it is puzzling that the conformational stability of this Qin et al.'s NTD construct was reported to very low, with most of the protein being unfolded at equilibrium, when Ubiquitin is a paradigm of a very stable, soluble protein [22, 23]. Indeed, Ubiquitin is behaves so well it is routinely chosen by NMR spectroscopists as an easy model protein to develop new pulse sequences. Here, we report the complete chemical shift assignments, high resolution NMR structure, conformational stability and backbone dynamics of the NTD.

5.3 Results

Production and purification of wt NTD and variants

As described in the Experimental Procedures section, we have successfully purified and produced a His-tagged N-terminus sequence (NTD) of TDP-43 that included residues 1 to 77 of this protein. The recombinant peptide was produced in bacteria and purified using a Ni-NTA resin (Qiagen) according to standard protocols.

Mass spectra

MALDI-TOF mass spectra of the NTD showed masses of 9923 – 9929 Da for the unlabeled protein (data not shown). These values are in excellent agreement with the theoretical mass of 9927 Da calculated for the sequence of this construct:
$M_{-11}R_{-10}G_{-9}SHHH_{-5}HHHG_{-1}$ S_0-

$M_1SEYI_5RVTED_{10}ENDEP_{15}IEIPS_{20}EDDGT_{25}VLLST_{30}VTAQF_{35}PGACG_{40}LRYRN_{45}P$
$VSQC_{50}MRGVR_{55}LVEGI_{60}LHAPD_{65}AGWGN_{70}LVYVV_{75}NY_{77}$, whose His-tag is shown in italics. The ^{13}C,^{15}N protein shows a mass peak at 10424.8 Da which is slightly lower than the calculated mass of 10480.9 Da for the ^{13}C,^{15}N-labeled protein. Based on this difference, the overall incorporation of ^{13}C and ^{15}N is calculated to be 89.9%.

NMR assignment of wild type NTD

The assignments are essentially complete; only the ^{15}N of prolines, the C' of residues preceding prolines, the side chain carbonyl carbons, the H-less C of aromatic residues; all the nuclei of the first two residues of the His tag and the ^{1}H and ^{15}N of the third residue of the His tag are missing. In the folded region of the protein; only the $^{13}C^{1}H_3\delta$ of M51 and ^{13}C' of Y77 (the C-terminal $-^{13}COO^-$) are missing. The assignments are unambiguous except for the usual cases involving geminal –H (in Lys, Arg, Met, etc), $-CH_3$ (in Val or Leu), and δ,ε signals in Tyr and Phe aromatic rings. The His-tag's six His share very similar chemical shift values. Two residues, R52 and G53, show two distinct $^{1}H^{15}N$ correlations which is evidence for conformational diversity. The NMR assignments have been deposited in the BMRB database under Access code **25675**. The ^{1}H-^{15}N spectra at 5ºC, 39 ºC and backbone and sidechain signals labeled at 25 ºC show the broad chemical shift dispersion typical of folded proteins and an absence of signals characteristic of denatured protein over 5 - 34 ºC (**Fig. 1**). A minor population of signals arising from unfolded protein is detected at 39 ºC. These signals disappear upon recooling.

The NTD adopts a well folded structure which resembles Axin 1 more than Ubiquitin

Secondary structure of the N-terminal domain consists of six β-strands; namely, β1: I5-T8, β2: I16-P19, β3: G40-R44, β4: R55-V57; β5 I60-H62 and β6: Y73-N76 and an α-helix composed of residues L28-Q34 (**Fig. 2 and Fig. 3**). These strands can be grouped into two β-sheets with the topologies β2 ***ap*** β1 ***p*** β6 ***ap*** β3 (sheet 1) and β4 ***ap*** β5 (sheet 2), where ***p*** and ***ap*** mean parallel and anti-parallel configurations, respectively. Whereas the topology of the first sheet is similar to that of Ubiquitin, the second sheet is a novel feature unique to TDP-43 NTD and to the reported structure of the C-terminal Dix domain of Axin 1, a scaffolding protein implicated in axonal growth and transport [24]. Strikingly, whereas their sequences are dissimilar, equivalent Cα atoms in the TDP-43 NTD and the Axin 1 Dix domain can be superimposed with an RMSD of only 2.2 Å (**Fig. 3F**), and this close structural similarity is characteristic of

homologous proteins [25]. Using the automatic program Dali, we found scores of proteins with significant structural similarity to the TDP-43 NTD. Expect for Axin 1, none has the same secondary structure topology, and none shared significant sequence similar to the TDP-43 NTD. The uniqueness of the NTD sequence and its distinct structure could explain the limited success in predicting its structure via fragment based *ab initio* modeling [21].

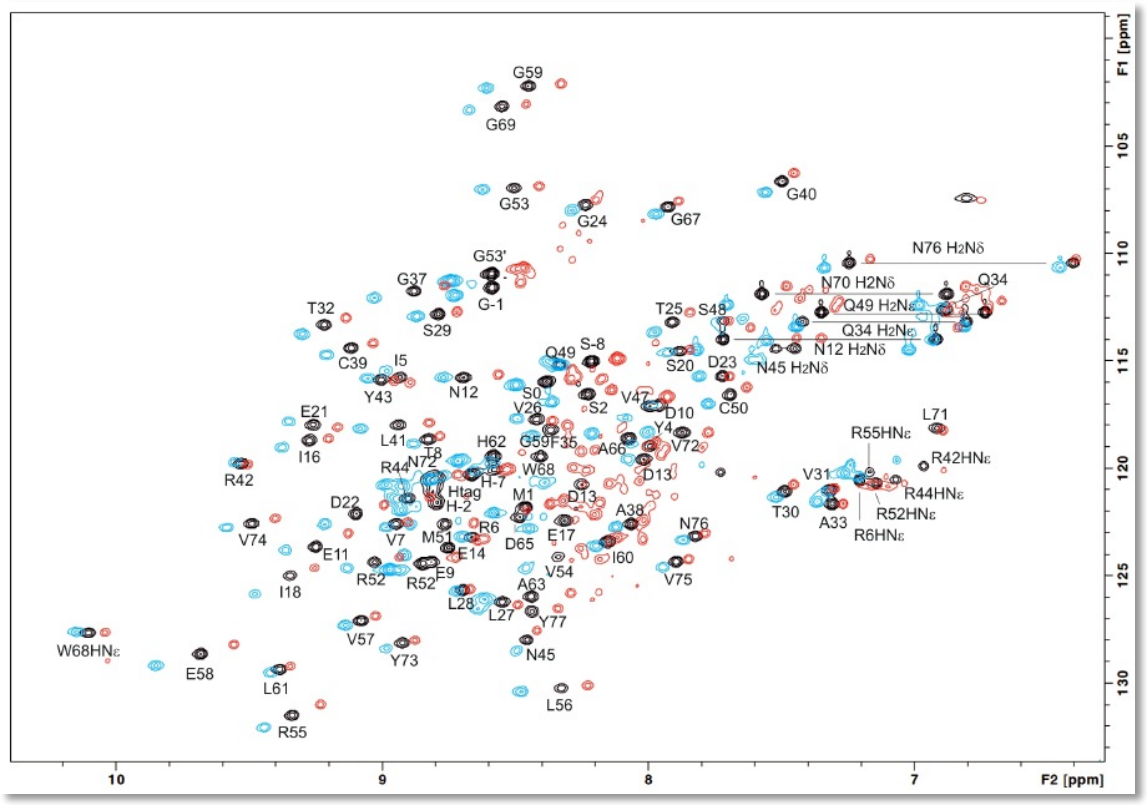


Figure 1: Assigned ^{1}H-^{15}N HSQC Spectra of the NTD at 5º, 25º and 39ºC. The ^{1}H-^{15}N HSQC spectra of the NTD at 5ºC (**blue**), 25 ºC (**black**) and 39 ºC (**red**) recorded in 1 mM TCEP, 3 mM Na/DAc(d_3), 1 mM NaN_3 in 85% MilliQ H_2O and 15% D_2O are shown. The ^{1}H chemical shifts and to a lesser extent those of ^{15}N, tend to decrease upon heating, but the changes are smaller for buried HNs forming H-bonds (for example, I5 and L61) and larger for solvent exposed HNs such as E58 and G69. Two distinct backbone HN signals were observed and assigned for R52 and G53. The second peak is labeled with a prime symbol. Whereas essentially no unfolded signals are seen at 5 ºC, 25 ºC and 34 ºC (not shown), a small population of resonances with denatured-like chemical shifts (between 8.3 – 7.95 ppm ^{1}H and 115 – 123 ppm ^{15}N) appear at 39 ºC. These signals disappear upon recooling. The side chain HNε of W68 and H_2Nε and H_2Nδ of Gln and Asn residues are labeled. The side chain HNε of Arg residues are folded and are also labeled.

Long-range backbone H-bonds between V26-L61, Y43-R52 and L41-V54 contribute the stability of the β-strands. In addition to these canonical structures, the residue segments E9- D13, E21-T25, F35-A38 and E58-G59 form well-defined turns that are stabilized by medium range backbone H-bonds. The NTD is stabilized by the burial of numerous nonpolar groups. Trp68 lies at the heart of the hydrophobic core

and makes multiple tertiary contacts. Its indole ring packs against Ala63 and Ile5, whose methyl groups experience intense ring current effects and are strongly shifted upfield. The Trp68 εNH is H-bonded to the O=C of Gly24 and shows slowed H/D exchange. Tyr43's phenyl moiety, also buried, packs edgewise against Trp 68's Hζ3 and Hη2 which are also shifted upfield. The side chains of Tyr 43 and Tyr 73, including their -OH groups, are buried in the hydrophobic core. The Tyr73 phenyl-O^{1}H signal is visible at 9.69 ppm in 1D ^{1}H and 2D NOESY spectra recorded in 85%H_2O/15% D_2O, but quickly disappears in 100% D_2O. Some resonance broadening and putatively distinct ^{1}Hδ / ^{1}Hδ' signals at 7.06 and 6.81 ppm reveal that the Tyr73's ring rotates relatively slowly on the ms timescale.

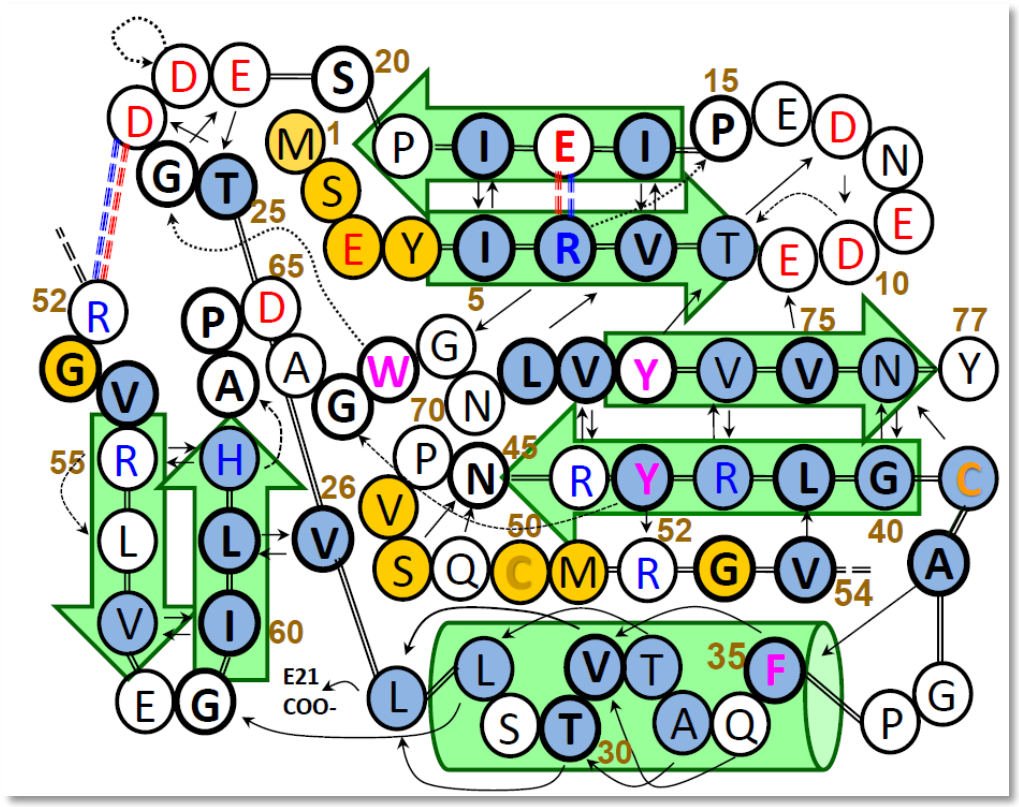


Figure 2: Summary of N-Terminal Domain Structure, Stabilizing Interactions and Dynamics. Transparent **green** arrows or a cylinder represent the β-strands and the α-helix, respectively. The double line marks the chain trace. Bold letters and thick circles indicate buried residues (SASA < 25%). **Blue** and **red** letters represent positively and negatively charged residues, respectively. **Blue** and **red** dashed lines indicate favorable charge - charge interactions that may form at 4.5 < pH < 10. Solid and dotted arrows represent backbone and sidechain H-bonds, respectively. W, F and Y residues are labeled **magenta**. Cys residues are labeled **amber**. Every fifth residue, and some other notable residues are numbered in **brown**. **Blue** shaded circles indicate residues whose HN group exchanges slowly. Orange shaded circles indicate residues whose HN group is dynamic (^{1}H^{15}N NOE ratio < 0.70).

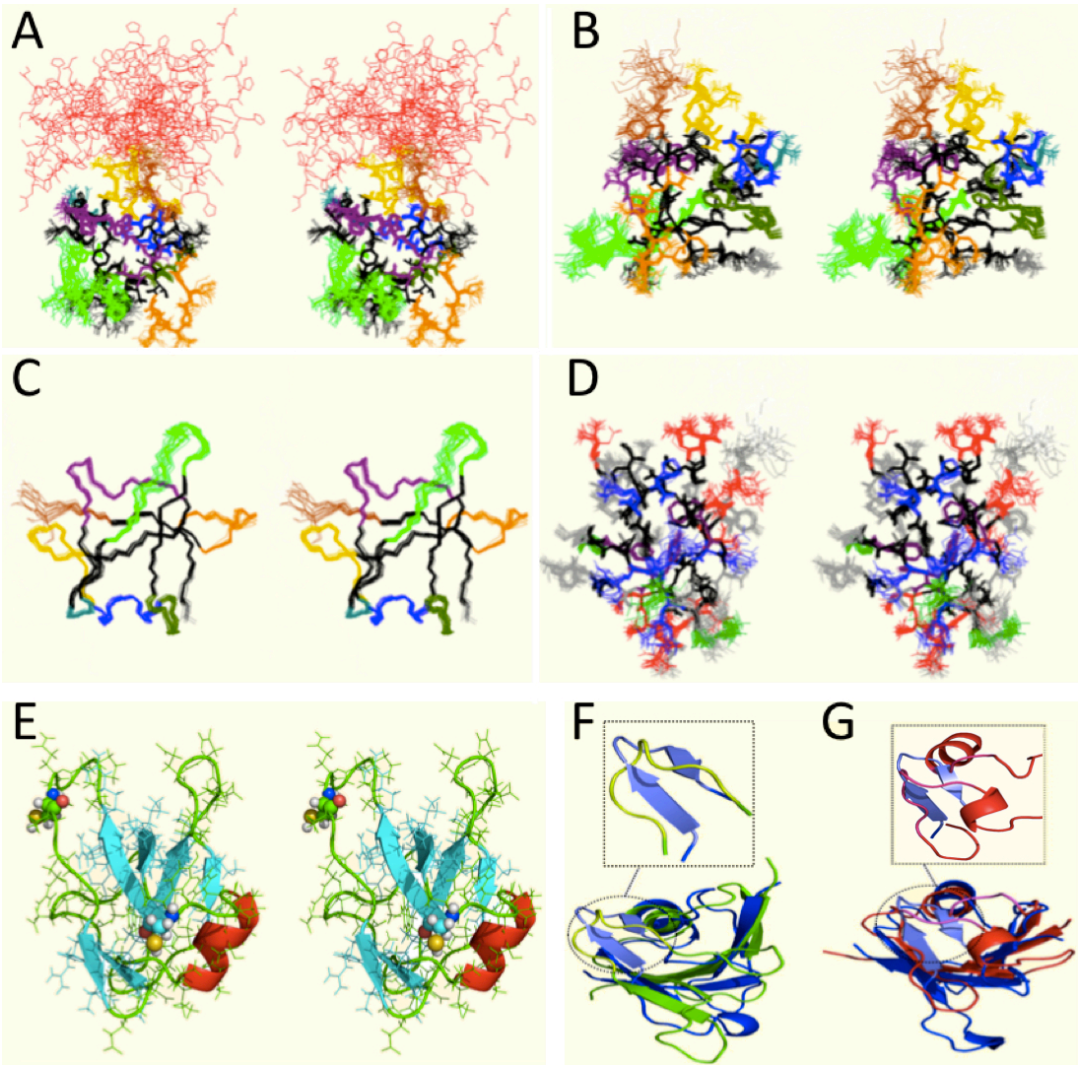


Figure 3: Structure of the N-terminal Domain. Cross-eyed stereo views of the structure of the NTD. In Panels **A**, **B** and **C**, residues in β-strands are colored **black**, the α-helix is shown in **blue** and residues in turn/loop segments are colored: **maroon** = residues 1-4; **orange** = 9-15; **golden** = 20-27; **green** = 35-39; **light green** = 45-54; **turquoise** = 58-59; **purple** = 63-72**;** **gray** =Y77. The His tag is shown in **red** in panel **A**. All heavy atoms are shown in panels **A**, **B** and **D**; the backbone only is shown in panel **C**. In panel **D** residues are colored as follows: cationic = **blue**, anionic = **red**, buried aliphatic = **black**; buried aromatic = **purple**; exposed hydrophobic = **green**; all others = **gray**. In panels **A**, **B**, **C** and **D** all 20 structures are shown. In panel **E**, the lowest energy structure is shown , each residue is numbered and the ribbon diagrams with the helix in **red** and strands in **cyan**. Cys residues are shown as spheres to highlight their far disposition within the NTD and their inability to form a disulfide bond in the folded, monomeric state. Panel **F** shows the structural superposition of the NTD (**blue**) and Axin 1 (**green**, pdb: 2D5G), whereas in panel **G** the NTD (in **blue**) is aligned with Ubiquitin (**red**, pdb: 1D3Z). The β-hairpin of the NTD is depicted in **slate blue** and magnified to show that this topological feature is present in the Axin 1 structure (indicated in **lime**), but not in ubiquitin (**pink**). These figures were prepared with the programs MOLMOL and PyMOL. The structures' orientation changes in each panel to afford the reader multiple perspectives.

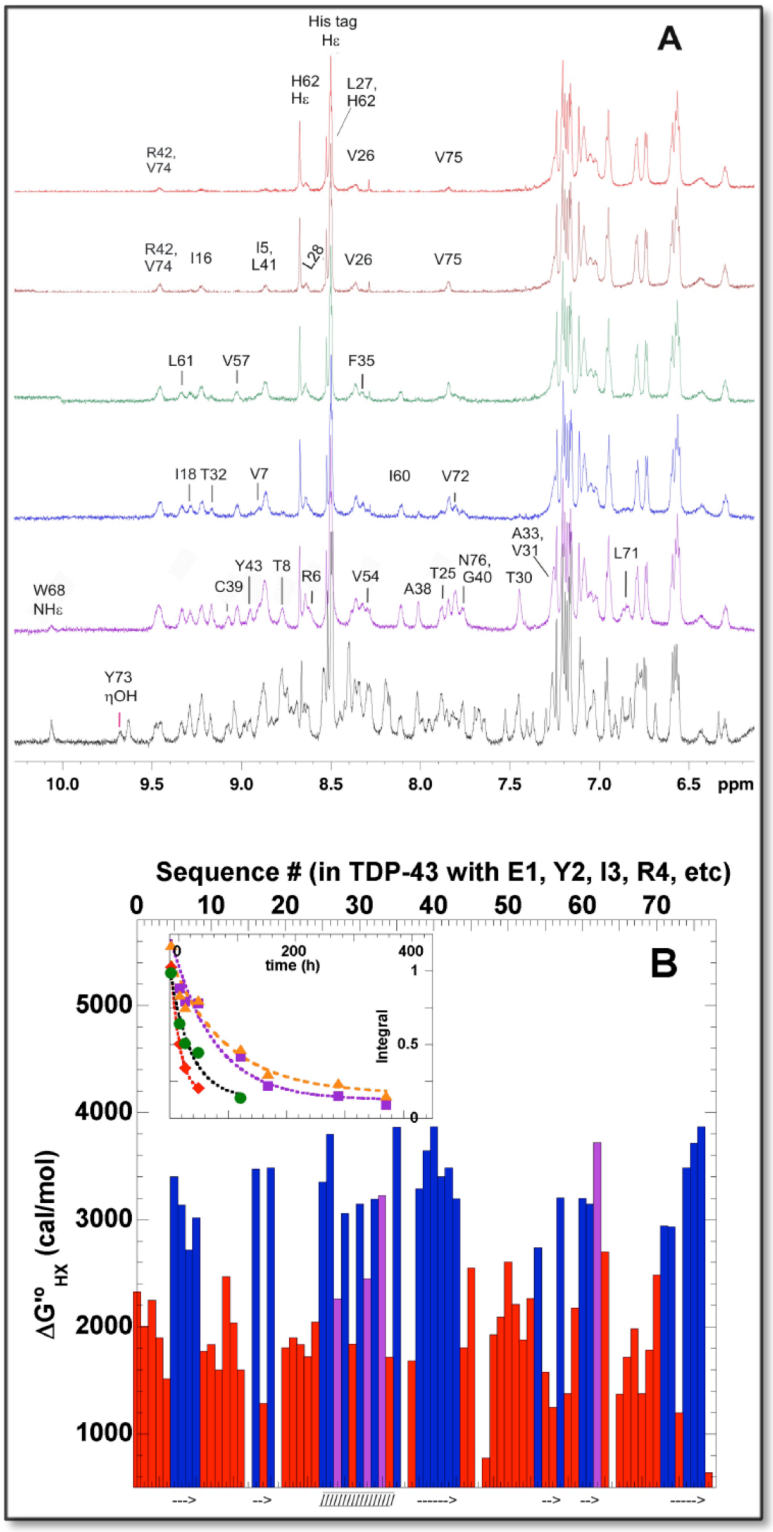
In addition to I3, G24, A63, W68, Y43 and Y73 mentioned above, residues R6, V7, I16, I18, S20, T25, V26, T30, V31, F35, A38, G40, L41, Y43, N45, G53, V54, G59, I60, L61, G67, L71, V72 and V75 are also predominately buried (%SASA < 25%). Regarding the Pro residues, P15 and P64 are also buried in the hydrophobic core. All the X-Pro peptide bonds are in the *trans* conformer as judged by the large difference between the $^{13}C\beta$ Pro and $^{13}C\gamma$ Pro chemical shifts [26]. Whereas most nonpolar side chains are buried within the hydrophobic core, P48, P58, V59 and M63 are exposed and could contribute to intermolecular interactions. P58 and V59 are rather close to the E9-D13 tight turn, which bears a high density of negative charge. The family of 20 refined structures has been deposited in the Protein Databank under access code: **2n4p**.

Since intradomain disulfide bond formation could contribute to the NTD's conformational stability, it is interesting to point out that the domain's two cysteine residues are located far apart from each other and would not be able to form a disulfide bond in the folded, monomeric state of this domain. As they are rather exposed, they may form disulfide bonds with other TDP-43 domains or with another NTD in a hypothetical dimer. Some studies have provided evidence that oxidized state of Cys residues in the RRM domains of full length TDP-43 contributes to aggregation [27, 28]. Here, both Cys are in the reduced state as determined on the basis of their $^{13}C\beta$ chemical shift.

H/D Exchange Reveals Many Slow Exchanging HN groups

The kinetics of hydrogen/deuterium exchange out of the NTD was followed by NMR spectroscopy (**Fig. 4A**). Out of a total of 72 backbone HNs in the domain, 31 exchange slowly. This is a fairly high ratio and is consistent with the elevated number of HN groups forming hydrogen bonds (**Fig. 2**). The indole HNε group of Trp 68, which is H-bonded to G24's carbonyl, also shows some protection against exchange. Some representative fits of a single exponential decay function to the peak integrals vs. time are shown in **Fig. 4B**. These rates are up to hundreds of times slower than the corresponding "intrinsic exchange rates" for a short, unstructured peptide with these sequence at pH* 3.8, 25ºC. Slow exchanging HN groups are observed in all elements of secondary structure, and the slowest are located in the internal β-strands or buried face of the α-helix, as is typically observed in other proteins such as Ubquitin [29] or Eosinophil Cationic Protein [30]. After exchange into D_2O was complete, the sample was exchanged back into 85% H_2O buffer and final 1D ^1H and 2D NOESY spectra were recorded. Although the signal to noise ratio of this spectrum was only fair, the resonances of S-8, the His tag, S0, E3, M12, E17, G24, G38, M51, R52, R55, E58, G59 and G69 could be unambiguously assigned. This confirms that these HNs have little or no protection against H/D exchange.

Figure 4: Residue-Level Stability of the NTD Measured by H/D Exchange



A. 1D ^{1}H NMR spectra of the downfield region of the NTD. The bottom spectrum (in **black**) was recorded in 85% H_20/15% D_2O and the subsequent spectra were recorded after 2h 10m=**purple**, 26h 25m=**blue**, 48h 35m= **green**; 168h 10m= **maroon** and 370h 10m=**red**. The non-exchanging side chain Hδ and Hε of His residues are also labeled.

B. Residue level conformational stability. **Red** bars represent upper limits of the conformational stability of fast exchanging residues. **Blue** bars stand for the measured stability of slow exchanging residues. **Purple** bars represent the minimum stability of HN groups which produce peaks in a 2D NOESY spectrum recorded in D_2O (and are therefore slow exchanging) but whose exchange kinetics could not be followed as they are overlapped with non-exchanging peaks in the 1D ^{1}H spectra. Arrows and the hashed parallelogram represent the position of β-strands and the α-helix, respectively. (*Inset*: H/D exchange of representative HN groups of the NTD. The peak integrals were normalized using non exchanging aromatic resonances using the TOPSPIN program. The curves show the fit of a single exponential decay function to each data set.)

The NTD has a conformational stability of 3.8 kcal/mol

The residue-level stabilities, calculated from the H/D exchange rates, is shown in **Fig. 4B**. The most stable residues are F47, V87 and N88 and their average stability is 3.8 ± 0.1 kcal·mol^{-1}. This value corresponds to the global conformational stability [51]. This is a fairly high conformational stability for a protein at such a low pH at 25ºC. For comparison, the conformational stabilities of remarkably stable proteins such as bovine Ubiquitin and RNase A are 6.9 and 6.7 kcal·mol^{-1}, respectively at pH 3.8, 25ºC in low ionic strength buffer [31, 32].

The N-terminal Domain is Stably Folded over a Wide Range of Temperatures and pH values

Although most spectra were recorded at pH 3.8 - 4.0 and 25 ºC, additional 1D ^{1}H and 2D ^{1}H NOESY spectra were recorded at pH 2.0 - 8.6 and 5 - 39 ºC. Native-like signals dominated the spectra under all these conditions including those of near physiological pH and temperature (**Fig. 5**). In addition, a series of 1D ^{1}H and 2D ^{1}H-^{15}N HSQC spectra were recorded at 5, 10, 15, 20, 25, 30, 34 and 39 ºC on the pH 3.9 sample (**Fig. 1**). Resonances that are characteristic of the native state dominate the whole temperature range. A small population of new peaks with chemical shift values typical of denatured proteins appears at 39 ºC, but disappears when the sample is cooled. These data provide strong evidence that the native domain is folded with little or no denatured structure under physiological conditions of pH and temperature. These spectra were analyzed to determine the temperature coefficients, the change of the ^{1}HN chemical shift as the temperature is varied ($\Delta \delta\ ^{1}H\ /\ \Delta T$) and these values are reported in **Table 2**. Small temperature coefficients ($|\ \Delta \delta\ ^{1}H\ /\ \Delta T\ | < 5$ ppB·ºC^{-1}) are associated with folded structure [33]. Here, all ^{1}HN showing low temperature factors are either H-bonded or buried or both, with the sole exception of D77. This excellent correlation is an additional validation of the structure reported here.

Fluorescence spectroscopy

The fluorescence emission spectrum of the NTD shows a maximum at 328 nm (**Fig. 6A**). This is typical for Trp indole moieties in hydrophobic environments and corroborates the burial of this group within the protein's hydrophobic core as determined by NMR methods (*vide supra*). Upon heating, the fluorescence emission decreases and above 44 ºC the emission maximum red-shifts reaching values over 350 nm which are indicative of solvent exposed Trp residues by 55 ºC (**Fig. 6B**). The variation of the wavelength of the emission maximum shows an apparent midpoint at 50 ºC at pH 4.0 and at 45 ºC at pH 7.0. By fitting a two state folded / unfolded model to the emission intensity data a T_M value of 42.1 ºC and a value of 3.0 kcal/mol for the conformational stability of the NTD at 25º, pH 4 can be calculated (**Fig. 6C**). The difference between the stability measured by fluorescence spectroscopy and H/D exchange can be ascribed to X-Pro *cis/trans* isomerization, which acts to increase the stability of the unfolded forms monitored by fluorescence but not those seen by H/D exchange [51].

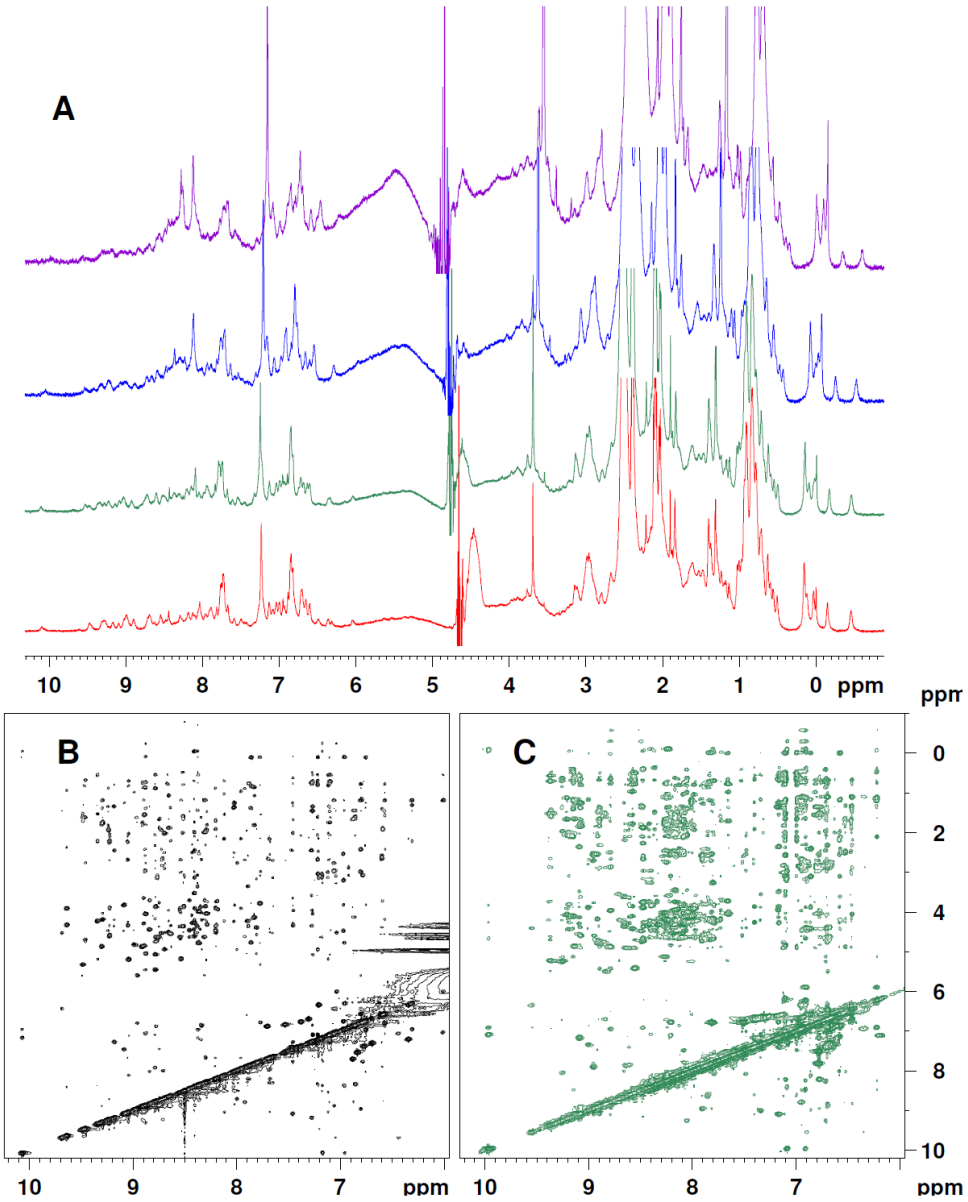


Figure 5: The NTD is Folded at Neutral pH. A. 1D ^{1}H NMR spectra of the NTD recorded at pH 7.15 and 5ºC (**purple**), 15ºC(**blue**), 25ºC(**green**) and 35ºC(**red**). Disperse ^{1}HN signals which are hallmarks of folded secondary structure and signals characteristic of the folded NTD hydrophobic core, such at the Ile5 δH at -0.40 to -0.60 ppm and the Trp68 Hζ3 at 6.3 ppm, are seen in all spectra. **B**. and **C**. Downfield region of the ^{1}H 2D NOESY spectra of the NTD at pH 3.75 (B, left in **black**) and at pH 7.15 (C, right in **green**). Scores of crosspeaks indicative of folded structure appear in both conditions. These spectra were recorded at 25ºC using a 150 ms mixing time.

Residue	Δ δ / Δ T (ppB / ºC)	Residue	Δ δ / Δ T (ppB / ºC)	Residue	Δ δ / Δ T (ppB / ºC)	Residue	Δ δ / Δ T (ppB / ºC)
M-11	n.d.[a]	P15	n.d.[b]	L41	-7.6	G67	-2.5
R-10	n.d.[a]	I16	-5.1	R42	-1.4	W68	-9.4
G-9	n.d.[a]	E17	-7.0	Y43	-2.9	G69	-2.5
S-8	-6.6	I18	-6.7	R44	-5.9	N70	n.d.[c]
H-7	-8.0	P19	n.d.[b]	N45	-2.2	L71	-1.4
H-6	-8.0	S20	-2.5	P46	n.d.[b]	V72	-6.7
H-5	-8.0	E21	-5.5	V47	-2.7	Y73	-3.2
H-4	-8.0	D22	-6.6	S48	-1.2	V74	-5.4
H-3	-8.0	D23	-3.4	Q49	-2.1	V75	-2.8
H-2	-7.3	G24	-9.6	C50	-4.3	N76	-2.0
G-1	-7.2	T25	-3.9	M51	-7.0	Y77	-7.1
S0	-5.8	V26	-3.7	R52	-5.7		
M1	-5.9	L27	-4.2	G53	-6.3		
S2	-6.6	L28	-1.6	V54	-6.0		
E3	-6.7	S29	-4.5	R55	-6.2		
Y4	-3.3	T30	-1.9	L56	-7.3		
I5	-2.6	V31	-0.8	V57	-3.3		
R6	-2.1	T32	-4.9	E58	-8.6		
V7	-2.4	A33	-3.0	G59	-8.2		
T8	-2.9	Q34	n.d.[c]	I60	-2.2		
E9	-6.7	F35	-3.6	L61	-2.2		
D10	-4.5	P36	n.d.[b]	H62	-3.6		
E11	-8.9	G37	-7.7	A63	-10.6		
N12	-2.2	A38	-7.7	P64	n.d.[b]		
D13	-3.0	C39	-5.2	D65	-4.8		
E14	-8.3	G40	-3.2	A66	-7.1		

Table 2: ^{1}H-^{15}N Temperature Coefficients: [a] n.d. = Not determined; peak not observed. [b]Prolines lack HN, [c]not determined due to peak overlap.

Contribution of Charged Residues to the Stability

The NTD contains an excess of negatively charged residues; there are only 1H + 0K + 5R = 6 cationic residues and 5D + 7E = 12 anionic residues. Most of the latter are concentrated in the turns connecting β-strand 1 to β-strand 2 and β-strand 2 to the α-helix and form two clusters of negative charge on the protein surface. To explore the contribution of electrostatic interactions to the conformational stability of the NTD, the pKas were predicted using the pKaPro algorithm. Most titratable groups were predicted to have pKas within 0.5 pH units of their intrinsic values (**Table 3**). In contrast, the pKas of Y43, H62 and Y73 are perturbed to favor the neutral forms due to unfavorable charge desolvation. E9 and E21 have moderately increased pKas as their charged forms are predicted to experience charge repulsion from D10 & D13 and D22 & D23, respectively. In other cases, interactions favor the charged form of the titratable groups. For example, D22's pKa is lowered as its charged form is favored due to H-bond formation. E14´s pKa is reduced and R6's pKa is raised due to a favorable electrostatic interaction between their charges. In a minority of structures, D23 is positioned to interact favorably with R52.

On the basis of the predicted pKa values, the pH dependence of the conformation stability ($\partial \Delta G / \partial$ pH) was calculated (**Figure 6D**). This calculation predicts that the NTD's conformational stability will be highest at acidic pH and will decrease as the pH is raised. This destabilization, which is substantial but insufficient to unfold the domain below pH 10, arises mainly from mutual repulsion of negatively charged carboxylate groups. The neutralization of H62 partially compensates as the loss its positive charge favors stability. At very alkaline pH (> 10) the titration of the two buried Tyr (Y43 and Y73) is strongly destabilizing.

These results contrast with those of Ubiquitin, which becomes more stable as the pH is increased [31]. This occurs because in Ubiquitin, charge – charge interactions are generally stabilizing at neutral pH [31]. However in the TDP-43 NTD, anionic residues tend to be clustered together and overall the charge – charge interactions are destabilizing at neutral pH.

Titratable Group	Predicted pKa	Intrinsic pKa	Comment
D10	4.28	3.80	
D13	4.04	3.80	
D22	**3.47**	**3.80**	**Negatively charged form stabilized by H-bonding to its own HN.**
D23	4.38	3.80	
D65	3.77	3.80	
E3	4.70	4.50	
E9	**5.23**	**4.50**	**Negatively charged form experiences repulsion from charged D10 & D13**
E11	4.62	4.50	
E14	4.58	4.50	
E17	**4.52**	**4.50**	**Negatively charged form interacts favorably with R6+ but is destabilized due to partial burial.**
E21	**5.03**	**4.50**	**Negatively charged form experiences repulsion from charged D22 & D23**
E58	4.58	4.50	
H62	**5.71**	**6.30**	**Charged form disfavored due to partial burial.**
C39	9.37	9.0	
C62	9.17	9.0	
Y4	10.39	10.0	
Y43	**12.69**	**10.0**	**Charged form strongly disfavored due to partial burial.**
Y73	**12.51**	**10.0**	**Charged form strongly disfavored due to partial burial.**
Y77	10.30	10.0	
R6	**12.80**	**12.50**	**Favorable charge-charge interaction with E17**
R42	12.34	12.50	
R44	12.42	12.50	
R52	12.26	12.50	
R55	12.26	12.50	

Table 3: Predicted pKa Values

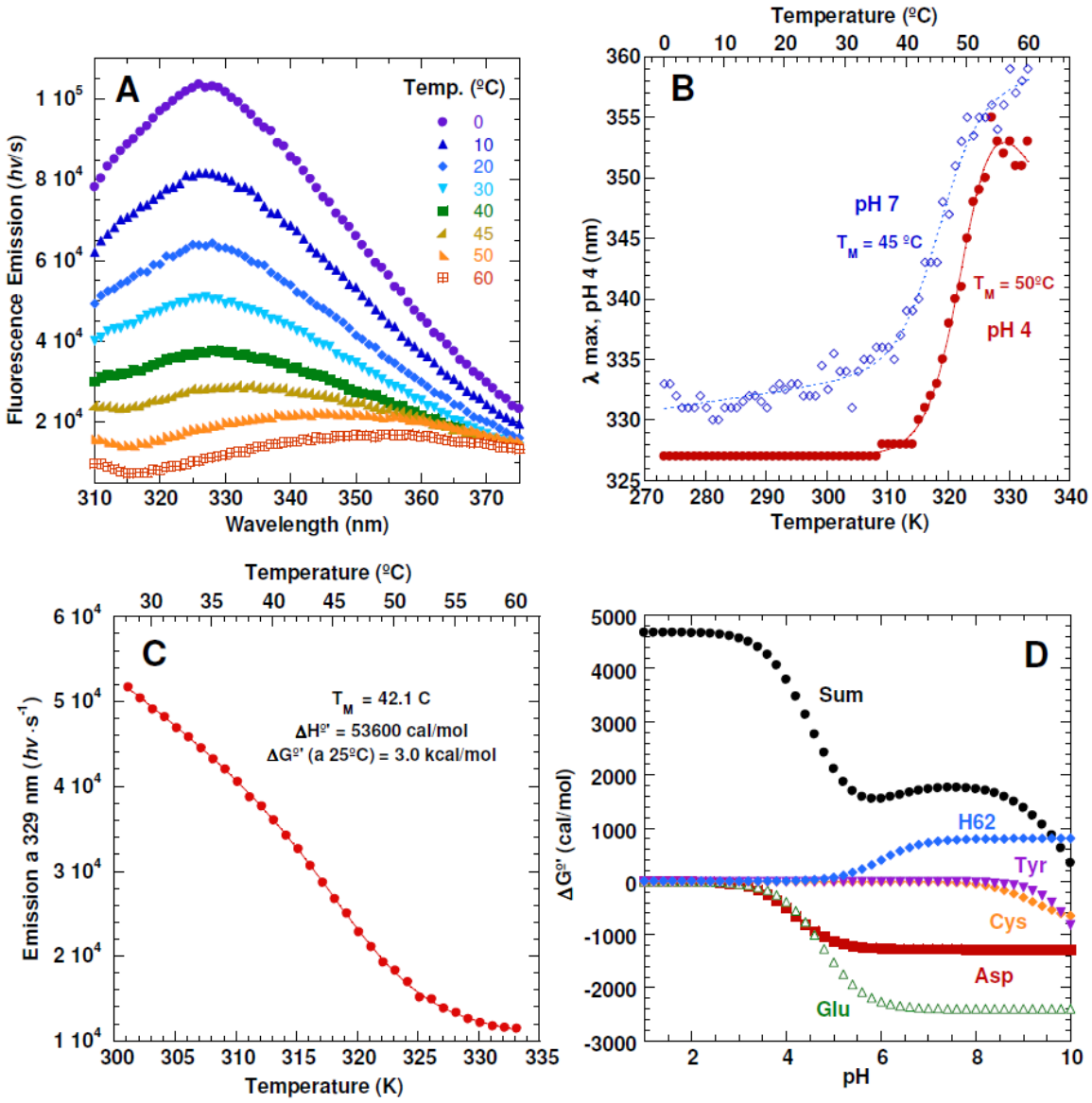


Figure 6: Fluorescence Spectra of NTD and Calculation of How pH Affects Stability. **A**. Fluorescence emission spectra of the NTD at pH 4 recorded at temperatures ranging from 0 to 60 ºC on a FluoroMax 4 instrument. The excitation wavelength was 280 nm. **B**. Variation in the wavelength of maximum emission upon heating. The apparent midpoints between values typical of buried and exposed Trp are indicated. **C**. The conformational stability of the NTD at pH 4 as determined by fitting a two-state denaturation model to the data, using the approach described in the Material and Methods section. **D**. The contribution of Asp (**red**), Glu (**green**), H62 (**blue**), Cys (**orange**) and buried Tyr 43 and Tyr 73 residues (**purple**) to the pH dependence of the conformational stability. These curves were calculated using the predicted values of the pKas in the folded and denatured states, as outlined in the Material and Methods section. The curve in **black** is the sum of the conformational stability determined by H/D exchange (3.8 kcal/mol at pH 4) plus the stabilizing/destabilizing contributions of the titratable groups as a function of pH.

NTD Backbone Dynamics Determined by NMR

The heteronuclear $\{^1H\}-^{15}N$ NOE and the longitudinal (R_1) and transverse (R_2) relaxation rates were measured for 64 of the 89 residues (**Fig. 7A,B,C**), including the His-tag. The longitudinal relaxation rates show little variation along the sequence of the NTD (**Fig. 7A**); their mean value is 1.70 ± 0.12 s^{-1}. The transverse relaxation rates are less homogeneous (**Fig. 7B**); they are smaller at the N-terminus and higher in the second β-strand. The mean R_2 value is 11.0 ± 2.3 s^{-1}, and was obtained from $R_1\rho$ measurements. The average NOE ratio were 0.76 ± 0.08 for residues 1-77 of the NTD, and many residues in secondary structural elements approach the value expected for a complete rigid HN vector at 25ºC and at this magnetic field strength (**Fig. 7C**). There were several residues with below average NOE ratios, in particular residues V47 – V54; this indicates flexibility on fast time scales (picoseconds to nanoseconds). The dynamic behavior of the N-terminal His-tag contrasts that of the NTD. Its average NOE ratio is much lower, only 0.39 ± 0.16, indicating that the His-tag is quite flexible.

The protein tumbles with an overall correlation time $\tau_m = 5.6 \pm 0.1$ ns, and behaves as an axial symmetric rotor of the oblate type (Dpar/Dper = 0.82 ± 0.12). The rotational diffusion properties have been also evaluated by hydrodynamic calculations and are in good agreement with experimental structure of a well-folded, monomeric, 77-residue domain (**Fig. 3A**). Thirty-three HNs have relaxation contributions from fast internal motions on ps-ns time scale and twelve are affected by slow exchange processes. Most of these residues are in loops or at the termini (**Fig. 7 D,E**). Most residues in secondary structural elements or turns possess values close to one of the general order parameter S^2, indicating that these residues are quite rigid (**Fig. 7F**). In contrast, other segments show lower S^2 values meaning that they are more mobile; the loop connecting β-strands 1 and 2, β-strand 2, which is exposed on the edge of the β-sheet and H-bonded only on one side, the termini and the loop composed of residues R44-G53.

The NTD Remains Folded After Six Months

To further test the robustness of the NTD, we recorded an additional spectra on the original sample used for the structure determination after six month of storage at 4 ºC. The 1D ^1H and 2D ^1H-^{13}C HSQC spectra reveal signals indicating that the hydrophobic core is natively packed and that the Cys residues are still reduced (data not shown). The 2D ^1H-^{15}N HSQC spectra on the sample freshly prepared and after six months are quite similar and show that the domain's backbone remains folded over this period (**Fig. 8**). These results show that the NTD is remarkably robust, in contrast to the metastable constructs achieved by Qin *et al* [21].

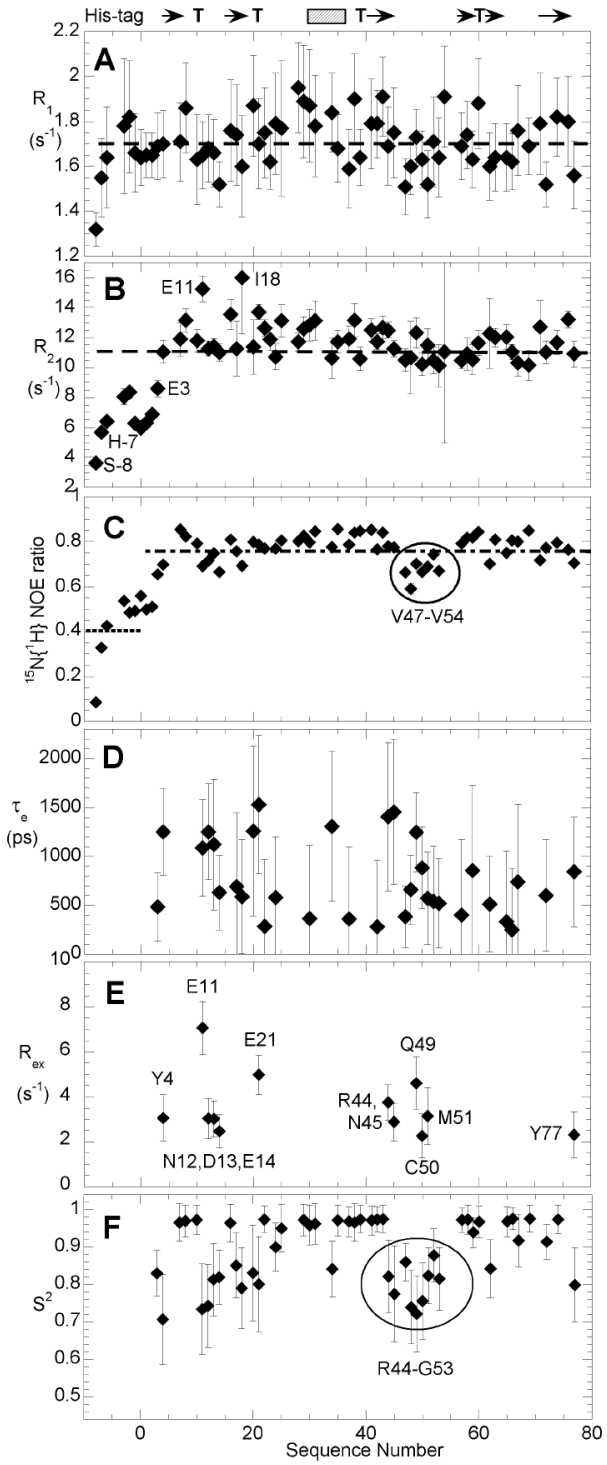


Figure 7: NMR Relaxation and NTD Dynamics. The positions of the elements of secondary structure are marked by arrows (β-strands), a hashed rectangle (α-helix) and "T" stands for well defined turns. Note that panels **A, B, C, D & E** share the same x-axis. The dashed line in **A** and **B** indicate the mean value.

A. Translational relaxation Rate (R_1).

B. Transversal relaxation rate (R_2).

C. $^{15}N\{^1H\}$ NOE-the dashed-dotted line is the average for the folded NTD and the dotted line is the mean of the flexible His-tag.

D. Effective time constants for fast internal motions, τ_e.

E. Contributions from slow conformational exchange, R_{ex}, to the dynamics.

F. The general order parameter, S^2.

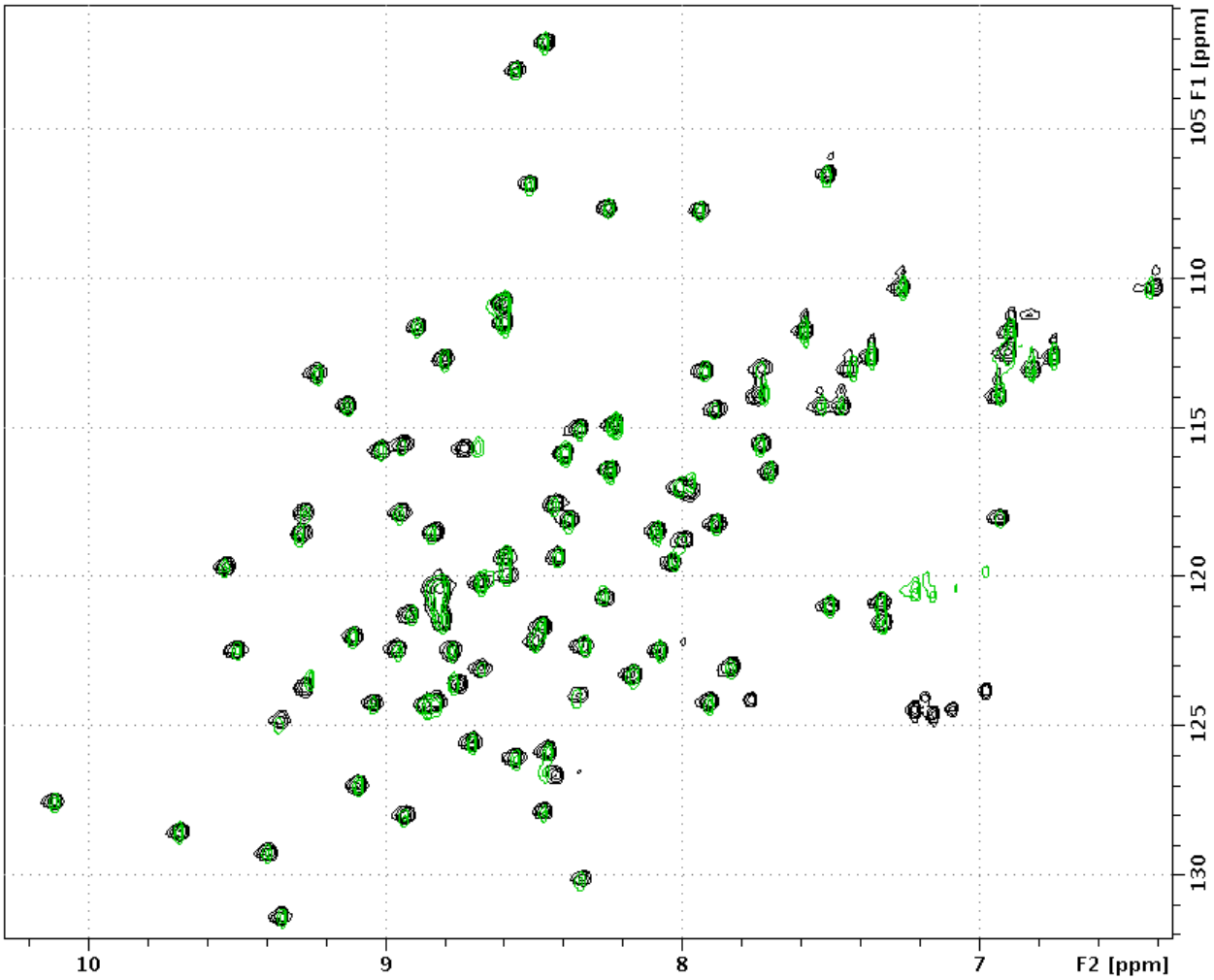


Figure 8: The NTD Remains Folded After Six Months. The ^{1}H-^{15}N HSQC spectra of the NTD at 25 ºC (**black**), previously shown with full assignment in Figure 1, and the same spectrum acquired after six months (green). All the ^{1}H-^{15}N signals overlap, which indicates that no structural changes have occurred after this long time. The position of the Arg side chain signals appear displaced in ^{15}N in the two spectra due to a difference in the ^{15}N spectral width.

5.4 Discussion

Several lines of evidence have recently highlighted the role played by TDP-43 N-terminal sequence in the aggregation of this protein and consequent loss-of-function effects at the RNA metabolic level. For this reason, it is extremely important to better understand its role in the pathological process. Previous deletion experiments, from some of us [20] and others [34], have suggested that structural elements could be involved in mediating TDP-43 loss of function effects following its aggregation within cells. For this reason, we decided to examine this connection in detail by starting with a detailed characterization of its structure.

Qin *et al.* studied three TDP-43 NTD constructs consisting of residues 1-80, 1-102, and 10-102, all with C-terminal His-tags, and obtained a low resolution structural model [21]. The structure of the NTD reported here corrects some of the structural elements that they found, confirms others and resolves many new structural elements. Qin *et al.* correctly identified most (four out of seven) of the secondary structural elements in the NTD; namely beta strands β1, β2, β3, β5 (our β6) and the α-helix. However, their predictions of the lengths of β3 and β5 (our β6) were too generous and the α-helix is significantly longer than they predicted it to be. The segment 49-52 was incorrectly identified as a β-strand by Qin *et al*. Although Q49 and C50 do donate H-bonds, they do so to N45, which lies just outside the 3rd beta strand. Thus, these H-bonds serve to close a turn structure. These H-bonds are too weak to afford detectable protection against H/D exchange (**Figure 5**). Expect for the mobile N-terminus, this segment is the most dynamic region of the domain, as revealed by both the $^{15}N\{^{1}H\}$NOE data of Qin *et al.* as well as our more complete investigation of the relaxation (**Figure 7**). The β-hairpin formed by residues 55 – 62 was not detected by Qin *et al*. This is an important element of the structure of the NTD which is stabilized by tertiary H-bonds with V26 and distinguishes the structure of the NTD from the Ubiquitin fold. This β-hairpin is rigid and the β-strand formed by residues 60-62 are linked to the rest of the protein by tertiary H-bonds and hydrophobic interactions (**Figures 2 and 3**). In addition, we have discovered and report four stable turn structures as well as the conformation of the side chains. The construct of the N-terminal domain studied here contains an N-terminal HexaHis tag for affinity purification and additional residues inserted for cloning: *MRGSHHHHHHGS*-M_1S_2-$E_3Y_4R_5$-etc. Three lines ofevidence, considered together, strongly suggest that these additional residues do not substantially perturb the conformation of the N-terminal domain. Firstly, the His tag and its adjacent residues show a very poor chemical shift dispersion and there are no NOEs detected between this segment and the folded NTD. In the structural calculations, this segment adopts a broad ensemble of conformers; the RMSD for His tag in the final set of 20 refined structures is 3.08 Å, which is much larger than the value of 0.67 Å determined for the rest of the molecule. Secondly, the His tag's ^{1}HN temperature coefficients are less than -6 ppB·ºC^{-1}; these values are typical of unstructured polypeptide [33]. Finally, the relaxation experiments show that this segments is mobile, whereas the folded NTD is chiefly rigid.

Another remarkable difference between our findings and those reported by Qin *et al.*, is that we find the NTD to be stably folded, with a conformational stability of almost four kcal/mol at 25ºC, pH 4, whereas they found the same domain to be in an equilibrium of folded and unfolded states under very similar experimental conditions in their 1-80 and 1-102 constructs, with the later being less stable. In considering what

could account for this remarkably different stability, it could be worth bearing in mind that Qin *et al.* placed their His-tag on the C-terminus of their NTD construct, whereas the His tag is positioned on the N-terminus of our construct. Our structural results do not support the proposal of Qin *et al.* that the folded/unfolded equilibrium is physiologically important. In fact, the conformational stability of our construction is such that even after six months the protein remained folded (**Figure 8**). In addition, the stabilization of the folded NTD conformer by nucleic acid binding observed by those researchers might result from relatively unspecific favorable electrostatic interactions. Residues 1-77 of the NTD have a net charge of about +5 at pH 4, and the longer 1-102 construct used by Qin *et al.* to study DNA binding carried a net charge of about +14 as the 81-102 segment contains 8 additional cationic residues (6K + 2R), only 5 anionic residues (4D + 1E) plus their C-terminal His-tag (6H). In our opinion, at neutral pH, the NTD would be more likely to interact with cationic partners through its two clusters of negatively charged residues.

The high sequence conservation of TDP-43's NTD throughout the animal kingdom [7] strongly suggests that it plays an important physiological role. Recently, some of us have found that the NTD is key for mediating the effective sequestering of TDP-43 into aggregative structures formed by C-terminal Q/N rich segments [20]. TDP-43 and FUS/TLS are members of the hnRNP family of proteins and both seem to form pathological aggregates implicated in ALS/FTLD as well as functional aggregates in RNA and stress granules [35]. FUS/TLS contains multiple repeats of a motif S/G-Y-S/G which has been proposed to be key for the reversible and kinase-controlled formation of a functional amyloid "hydrogel" which play roles in RNA regulation and transport [36]. Similar multiple S/G-Y-S/G -like motifs are also found in the C-terminal portion of TDP-43 and flank its Q/N rich motif. Although TDP-43's N-terminal domain appears to be unique among hnRNP proteins, it may work like the domains or accessory proteins which govern amyloidogenesis in functional amyloids, including curlins [37], Sup35 [38] and CPEB/Orb2A [39]. Therefore, the high-resolution structure of the NTD of TDP-43 may represent the first example of an emerging class of protein domains that regulate functional amyloid formation. For this reason, we expect that this structure and the chemical shift assignments reported here will be valuable tools for studying these interactions between the NTD and C-terminal aggregative segments of TDP-43.

5.5 Experimental procedures

Sample Preparation

The sequence coding for TDP-43 N-terminus residues (1-77) was cloned in the pQE30-6xHis plasmid using the BamHI-HindIII restriction enzymes according to standard protocols. The insert was confirmed by sequencing and the plasmid was used to transform M15 bacteria that were grown on Kana/Amp resistant plates. For expression, the bacteria were then grown in M9 media supplemented with D-Glucose (U-13C, 99%, Cambridge Isotope lab, Inc) at 0.3% v/v and Ammonium-15-Chloride (Sigma) at 0.1% v/v. Protein expression was induced overnight with 1 mM IPTG at 30°C and the NTD peptide was purified using Ni-NTA Agarose Qiagen Resin in the presence of Complete PI (Roche) according to manufacturer's instructions. The protein was eluted from the resin using 250 mM imidazole and each fraction was loaded on a 15% SDS-PAGE before staining with Colloidal Coomassie Blue (**Sup. Fig. 1**).

The protein was then frozen and shipped to Spain. At the IQFR in Madrid, the protein samples were allowed to thaw. Then urea was added (to a final concentration of 1 M) and the pH was lowered from about 8.5 to 3.8 by adding HCl and CH_3COOH. The presence of urea helps dissolve the protein and prevents the formation of aggregates when the pH is near the pI of the NTD (calculated to be 4.4 for the NTD and 5.9 for this construct with its hexaHis tag). Next, the sample was concentrated to 1 mL using a Amicon ultrafiltration device with a 3 kDa cutoff and applied to a GE Biosystems PD-10 gel filtration column that had been preequilibrated with buffer consisting of 85% H_2O MilliQ, 15% D_2O, 1.0 mM TCEP, as the reducing agent, 1.0 mM NaN_3, to prevent microbial contamination, and 3.0 mM CD_3COOD/ $CD_3COO^-Na^+$ to buffer the pH at 3.8. Half milliliter fractions were collected and the NTD-containing fractions were identified by UV absorbance or (more conveniently) fluorescence spectroscopy. MALDI-TOF mass spectrometry was performed to confirm the identity of the sample. The NTD-containing fractions were concentrated to a final volume of approximately 0.25 mL and put in a water matched Shigemi NMR tube. NMR spectra were recorded on a Bruker 800 MHz Advance NMR spectrometer, equipped with a ^1H,^{13}C,^{15}N cryoprobe and z-gradients. Whereas most experiments were performed at pH 3.8, 25ºC, some NMR spectra were recorded over a range of pH (from 2 to 8.5) and temperatures (from 5ºC to 39 ºC). Additional spectra were recorded on the buffer without protein, either alone or with excess hydrogen peroxide. By analyzing these spectra, the chemical shift values of the methylene groups of TCEP were found to be multiplets centered at 2.65 and 2.43 ppm (reduced form) and 2.50 and 2.08 ppm (oxidized form) at pH 3.8 and 25ºC. These values and 1D ^1H NMR spectra were utilized to confirm the presence of reduced TCEP in sample during the course of acquiring 2D and 3D NMR spectra.

Chemical Shift Referencing

As preliminary spectra revealed that the NTD has NMR signals near 0 ppm, the standard NMR chemical shift reference compound DSS was not included in the majority of the samples. Instead, the chemical shift reference was set by measuring the DSS signal in buffer. A series of 1D ^1H experiments with this reference sample were recorded at several different temperatures in order to determine the correct reference value for the experiments used to measure the ^1HN temperature coefficients. The ^{13}C and ^{15}N chemical shift reference values were calculated from the ^1H reference by multiplying by nuclei's gyromagnetic ratios [40].

Chemical shift assignment

The assignment of TDP-43 was carried out successfully using a standard 3D approach based on sequentially connectivities between ^1H, ^{15}N nuclei and the $^{13}C\alpha$ and $^{13}C\beta$ of the same residue and the preceding one. This assignment process was corroborated by 2D ^1H-^{15}N HSQC spectra which are ^{13}C filtered according to the side chain structure as well as spectra which specifically detect residues proceeding or following proline. The backbone assignment process was carried out automatically using MARS [41] as well as manually using by two different operators with the aid of the programs Sparky [42] and CcpNmr [43]. Only a few differences were observed and these could be resolved by careful study and analysis of additional spectra. Additional verification was obtained by registering and analysis of a (4,2)D $^1H^{15}N$-$^1H^{15}N$ spectrum. Side chain chemical shifts were assigned by studying 3D HCCCNH, CCCCNH "TOSCY-HSQC") type spectra. These assignments were checked and a Tyr -O^1H signal was assigned by analysis of 1D ^1H and 2D ^1H NOESY spectra recorded in

85%/15% H_2O/D_2O or 100% D_2O. The complete list of the NMR spectra used to assign and obtain structural constraints is given in the **Appendix I**.

Table 1: Structure Calculation and Validation Statistics

NOE Distances and Dihedral Constraints	
# short-range distances ($\|i-j\| \leq 1$)	511
# medium-range distances ($1 < \|i-j\| < 5$)	123
# long-range distances ($\|i-j\| \geq 5$)	424
# restraints from H-Bonds	26
# angular restraints (ϕ,ψ)	60
# total restraints	1058 + 26 + 60 = 1144
#restrictions / residue	13.74
Cyana target function value	1.09
AMBER Refinement Statistics	
RMSD from experimental restraints (NOE) (min, mean, max)	0.07, 0.13, 0.53
RMSD from dihedral restraints (TALOS) (min, mean, max)	2.54, 4.81, 7.06
RMSD bond lengths from ideal geometry (Å)	0.01
RMSD angles from ideal geometry (Å)	2.68
Averaged pairwaise RMSD (Å) (20 final structures)	
Mean Pairwise RMSD for the His tag (residues -9 - 0)	
Backbone	3.08 ± 0.78 Å
All heavy atoms	4.82 ± 0.72 Å
Mean Pairwise RMSD (residues 1-77, this excludes the His-tag) for the family of 20 structures	
Backbone	0.67 ± 0.18Å
All heavy atoms	1.38 ± 0.18Å
Mean Pairwise RMSD for Secondary Structure Elements (residues: 5-8, 16-19, 28-35, 40-44, 55-57, 60-62, 72-76)	
Backbone	0.23 ± 0.05 Å

Structure Calculation

The solution structure of the N-terminal domain TDP-43 (1-77) was determined on the basis of backbone torsional angle restrictions derived from NMR parameters NOE distance constraints. Restraints for φ and ψ dihedral angles were derived from $^{1}H_{\alpha}$ chemical shifts using the program TALOS+ [44]. Distance constraints were obtained from 3D NOESY- ^{13}C-^{1}H HSQC and 2D ^{1}H NOESY spectra. The later was acquired in 100% D_2O in order to obtain the most restrictive restraints. Structures were calculated following a three-step procedure. First, automatic NOE assignment of the spectra was carried out as implemented in the program CYANA v2.1 [45]. This protocol performs seven cycles of automated NOE assignment and structure calculation and a final annealing using the list of restraints obtained in the last cycle. This set of distance constraints was checked by visual inspection of the spectra. In a second step, we add a number of H-bond restrictions that were obtained from an H/D exchange experiment (*vide infra*) and we performed a standard simulated annealing to obtain 100 conformers from which those 20 with lowest target function values were selected. Finally, we used the AMBER9 [46] software to refine this ensemble of structures. To this end, 2000 steps of energy minimization with implicit solvation were performed over the selected conformers. The structural statistics data for the final ensembles of 20 structures obtained for TDP-43 (1-77) are provided in **Table 1**. The quality of these final structures was assessed using the Protein Data Bank validation server (http://validate.rcsb.org).

Structure Comparison

The NTD structure was manually compared to known folds in the SCOP structural database [47], and the whole PDB was surveyed automatically using the Dali server [48].

Hydrogen/deuterium Exchange

In order to measure the hydrogen/deuterium (H/D) exchange kinetics, a PD-10 column was utilized to exchange the NTD into 3.0 mM acetic acid/sodium acetate, 1.0 mM sodium azide, 1.0 mM TCEP buffer prepared with 100% D_2O. Then, a series of eleven 1D ^{1}H spectra were recorded to follow exchange over the course of two weeks. After approximately twelve hours of exchange, a 2D ^{1}H NOESY spectrum was recorded. The analysis of this spectrum served to corroborate the identity of the slow exchanging HN groups, especially those which are overlapped with other exchanging peaks or non-exchanging aromatic resonances. Between measurements, the sample was incubated in a water bath at 25 ºC. The temperature and water level of the bath was checked every two or three days and the temperature variation was less than ± ½ ºC. The HN peaks and non exchanging aromatic signals, which were used as internal controls, were integrated using Topspin 2.1 and a single exponential decay equation was fit to the data to determine the observed exchange rate for each residue. By comparison to the intrinsic exchange rates calculated using the parameters of [49], the conformational stability for each residue and the global stability were determined by assuming the EXII exchange mechanism [50] (which holds for low pH conditions) and following the approach of Pace and co-workers [51]. After H/D exchange was complete, the sample was back-exchanged into 85% H_2O /15% D_2O buffer and a 2D ^{1}H NOESY spectrum was recorded to corroborate the identity of the fast exchanging HN groups.

Fluorescence spectroscopy

A Fluoromax 4 spectroflourimeter (Jobin/Yvon) equipped with a Peltier module for temperature control was used to record fluorescence spectra on the NTD dissolved in the same buffer (85% H_2O/15% D_2O) employed for NMR spectroscopy. The instrument's excitation and emission wavelengths were calibrated using the Xe emission line at 467 nm and the water Raman signal, respectively. The excitation wavelength was 280 nm and the emission was recorded over from 300 – 400 nm with a 2 $ns \cdot s^{-1}$ scan speed at 25 ºC. For monitoring thermal denaturation, the fluorescence spectra were recorded over 1 ºC intervals. Once the temperature was raised, the sample was allowed to equilibrate for one minute prior to the acquisition of a fluorescence spectrum over the wavelength range of 310 – 375 nm. The slit widths were 3.0 nm for excitation and emission in all spectra. A two-state denaturation model with linearly sloping pre- and post-transition baselines was fit to the data to obtain values for T_M, the temperature midpoint of thermal denaturation, as well as ΔH_M and ΔS_M, the enthalpy and entropy changes for denaturation, respectively. The Gibbs-Helmholtz equation was employed to estimate the conformational free energy at 25ºC, $\Delta G°'(25ºC)$:

$$\Delta G°' (25ºC) = \Delta H°'_M [1\text{-}298.15 \text{ K} / T_M] – \Delta Cp[(T_M\text{-}298.15) – 298.15*\ln(298.15/T_M)] \quad (1)$$

using a value of 1132 $cal \cdot M^{-1} \cdot K^{-1}$ for ΔCp, the heat capacity change for denaturation, which was estimated using the empirical equations of Pace and coworkers [52] for a 77 residue globular protein.

pH Dependence of the Conformational Stability

Using the refined structures of the NTD as input, we utilized the program PropKa [53] to estimate pKa values for the domain's titratable groups in the folded state ($K_{a,j,N}$). These values, and standard values for the pKas of the same groups in the denaturated state ($K_{a,j,D}$) were then used to calculate the dependence of the conformational stability on pH ($\partial \Delta G / \partial$ pH) as previously described [54] using the equation reported by Tanford [55]:

$$\frac{\partial \Delta G°'}{\partial pH} = -RT \ln \left[\frac{\prod_{j=1}^{n} (1 + K_{a,j,D} / a\, H^+)}{\prod_{j=1}^{n} (1 + K_{a,j,N} / a\, H^+)} \right] \qquad (2)$$

where aH^+ is the activity of the hydrogen ion, which is assumed to be equal to the hydrogen ion concentration: $aH^+ \approx [H^+] = 10^{(-pH)}$. This is an excellent approximation for solutions of low ionic strength above pH 2.

Dynamics: ^{15}N Relaxation

All NMR experiments were acquired at 800MHz, pH 3.8 and 25 ºC. A set of ^1H-^{15}N correlation spectra were registered to measure the ^{15}N longitudinal (R_1), off resonance rotating frame ($R_1\rho$), and transverse (R_2) relaxation rates. Three sets of 10 spectra with relaxation delays ranging from 20 to 1600, 8 to 200, and 0 to 285 ms were recorded to determine R_1, $R_1\rho$, and R_2, respectively. The heteronuclear $^{15}N\{^1H\}$ NOEs of individual backbone HN groups were measured as the ratio of spectra recorded with and without saturation in an interleaved fashion. A long recycling delay of 10 s was employed to ensure the maximal development of NOEs before acquisition and to allow solvent relaxation, thus avoiding transfer of saturation to the most exposed amide protons of the protein between scans [56]. The principal components of the TDP-43(1-77) NTD inertia tensor were calculated with Pdbinertia (A. G. Palmer, III, Columbia University, New York, NY) using the lowest energy structure. Heteronuclear R_2 values

were derived from $R1\rho$ measurements. We estimated the overall correlation time from the ratio of the mean values of R_1 and R_2. The values for R_1 and R_2 were calculated from a subset of residues with little internal motion and no significant exchange broadening. This subset excluded residues with NOEs of <0.65 and also residues with R_2 values lower than the average minus one standard deviation, unless their corresponding R_1 values were larger than the average plus one standard deviation [57]. The diffusion tensor, which describes rotational diffusion anisotropy, was determined by two approaches [58, 59] as implemented in the r2r1_diffusion and Quadric_diffusion programs by A.G. Palmer (A. G. Palmer, III, Columbia University). The ^{15}N relaxation was analyzed assuming dipolar coupling with the directly attached proton (with a bond length of 1.02 Å), and a contribution from the ^{15}N chemical shift anisotropy evaluated as −160 ppm. Relaxation data were fitted to the Lipari and Szabo model [60] using FAST-Modelfree [61], which interfaces with MODELFREE version 4.2 [62]. Five models of internal motion were evaluated for each amide $^{1}H-^{15}N$ pair: (i) S^2, (ii) S^2 and τ_e, (iii) S^2 and R_{ex}, (iv) S^2, τ_e, and R_{ex}, and (v) S_f^2, S^2, and τ_e, where S^2 is the generalized order parameter, τ_e is the effective internal correlation time, $R_{ex}$is the exchange contribution to transverse relaxation, and S_f^2 is related to the amplitude of the fast internal motions.

5.6 Acknowledgments

This work was supported by Grants CTQ2010-21567-C02-02 (FPI Fellowship to MMG), SAF2013-49179-C2-2-R (DVL), EU Joint Programme-Neurodegenerative Diseases (FPND) grants RiMod-FTD, Ministerio della Sanita', Italia (EB) and AC14/00037 (DVL), AriSLA TARMA (FB) and Theirry Latran Foundation REHNPALS (EB).

5.7 References

1. Ou SHI, Wu F, Harrich D, García-Martínez LF & Gaynor RB (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Vir* **69**, 3584-3596
2. Buratti E, Dörk T, Zuccato E, Pagani F, Romano M & Baralle FE (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* **20**, 1774-1784
3. Buratti E & Baralle FE (2012) TDP-43: Gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem Sci* **37**, 237-247
4. Neumann M, Sampathu DM & Kwong LK (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133
5. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, Mann D, Tsuchiya K, Yoshida M, Hashizume Y & Oda T (2006) TDP-43 is a component

of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun* **351**, 602-611

6. Josephs KA, Whitwell JL, Weigand SD, Murray ME, Tosakulwong N, Liesinger AM, Petrucelli L, Sengem ML, Knopman DS, Boeve BF, Ivnik RJ, Smith GE, Jack CRJ, Parisi JE, Petersen RC & Dickson DW (2014) TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol* **127**, 811-824
7. Ayala YM, Pantano S, D'Ambrogio A, Buratti E, Brindisi A, Marchetti C, Romano M & Baralle FE (2005) Human, Drosophila and C. elegans TDP-43: Nucleic acid binding properties and splicing regulatory function. *J Mol Biol* **348**, 575-588
8. Guo W, Chen Y, Zhou X, Kar A, Ray P, Chen X, Rao EJ, Yang M, Ye H, Zhu L, Liu J, Xu M, Yang Y, Wang C, Zhang D, Bigio EH, Mesulam M, Shen Y, Xu Q, Fushimi K & Wu JY (2011) An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat Struct Mol Biol* **18**, 822-830
9. Robinson JL, Geser F, Stieber A, Umoh M, Kwong LK, vanDeerlin VM, Lee VMY & Trojanowshi JQ (2013) TDP-43 skeins show properteis of amyloid in a subset of ALS cases. *Acta Neuropathol* **125**, 121-131
10. Bigio EH, Wu JY, Deng HX, Bit-Ivan EN, Mao Q, Ganti R, Peterson M, Siddique N, Ceula C, Siddique T & Mesulam M (2013) Inclusions in frontotemporal lobar degeneration with TDP-43 proteinopathy (FTLD-TDP) and amyotrophic lateral sclerosis (ALS), but not FTLD with FUS proteinopathy (FTLD-FUS) have properties of amyloid. *Acta Neuropathol* **125**, 463-465
11. Chen AKH, Lin RYY, Hsieh EZJ, Pu PH, Chen RPY, Liao TY, Chen WW, Wang CH & Tauang JJT (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J Am Chem Soc* **132**, 1186-1187
12. Saini A & Chauhan VS (2011) Delineation of the core aggregation sequences of TDP-43 C-terminal fragment. *ChemBioChem* **12**, 2495-2501
13. Jiang LL, Che MX, Zhao J, Zhou CJ, Xie MY, Li HY, He JH & Hu HY (2013) Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J Biol Chem* **288**, 19614-19624
14. Saini A & Chauhan VS (2014) Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir* **30**, 3845-3856
15. Budini M, Buratti E, Stuani C, Guarnaccia C, Romano V, De Conti L & Baralle FE (2012) Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. *J Biol Chem* **287,** 7512-7525
16. Mompean M, Buratti E, Guarnaccia C, Brito RMM, Chakrabartty A, Baralle FE & Laurents DV (2014) Structural characteriation of the minimal segment of TDP-43 competent for aggregation. *Arch Biochem Biophys* **545**, 53-62
17. Mompeán M, Hervás R, Xu Y, Tran TH, Guarnaccia C, Buratti E, Baralle FE, Tong L, Carrión-Vázquez M, McDermott AE & Laurents DV (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J Phys Chem Lett* **6**, 2608-2615
18. He RY, Huan YC, Chiang CW, Tsai YJ, Ye TJ. Gao HD, Wu CY, Lee HM & Huang JJT (2015) Characterization and real-time imaging of the FTLD-related

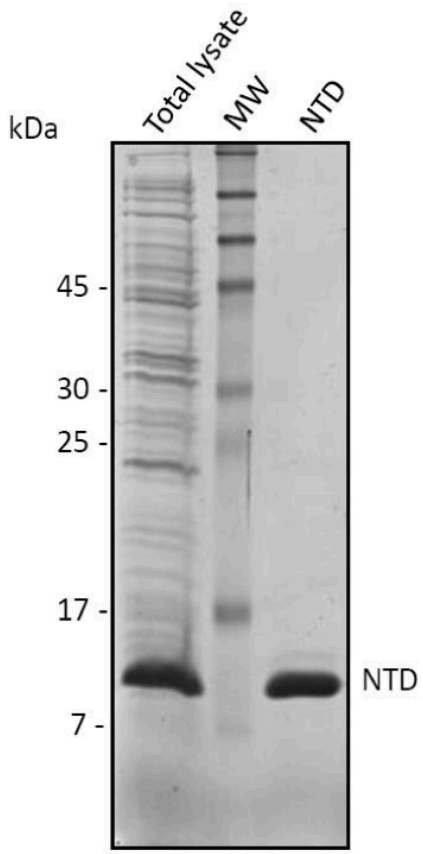
protein aggregation induced by amyloidogenic peptides. *Chem Comm* **51**, 8652-8655
19. Budini M, Romano V, Quadri Z, Buratti E & Baralle FE (2015) TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum Mol Genet* **24**, 9-20
20. Romano V, Quadri Z, Baralle FE & Buratti E (2015) The structural integrity of TDP-43 N-terminus is required for efficient aggregate entrapment and consequent loss of protein function. *Prion* **9**, 1-9
21. Qin H, Lim LZ, Wei Y & Song J (2014) TDP-43 N-terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc Natl Acad Sci USA* **111**, 18619-18624
22. Briggs MS & Roder H (1992) Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci USA.* **89**, 2017-2021
23. Lazar GA, Desjarlais JR & Handel TM (1997) De novo design of the hydrophobic core of ubquitin. *Protein Sci* **6**, 1167-1178
24. Hida T, Nakamura F, Usui H, Takeuchi K, Yamashita N & Goshima Y (2015) Semaphorin3A-induced axonal transport mediated through phosphorylation of Axin-1 by GSK3b. *Brain Res* **1598**, 46-56
25. Subbiah S, Laurents DV & Levitt M. (1993) Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Curr Biol* **3**, 141-148
26. Schubert M, Labudde D, Oschkinat H & Schmieder P (2002) A software tool for the prediction of Xaa-Pro peptide bond conformation in proteins based on 13C chemical shift statistics. *J Biomol NMR* **24**, 149-154
27. Cohen TJ, Hwang AW, Unger T, Trojanowski JQ & Lee VMY (2011) Redox signalling directly regulates TDP-43 via cystein oxidation and disulphide cross-linking. *EMBO J* **1-12**, 1-12
28. Chung CK, Chiang MH, Toh EK, Chang CF & Huang TH (2013) Molecular mechanism of oxidation-induced TDP-43 RRM1 aggregation and loss of function. *FEBS Lett* **587**, 575-582
29. Schanda P, Forge V & Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci USA* **104**, 11257-11262
30. Laurents DV, Bruix M, Jiménez MA, Santoro J, Boix E, Moussaoui M, Nogués M & Rico M (2009) The (1)H, (13)C, (15)N resonance assignment, solution structure and residue level stability of eosinophil cationic protein/RNase 3 determined by NMR spectroscopy. *Biopolymers* **91**, 1018-1028
31. Ibarra-Molero B, Loladze VV, Makhatadze GI & Sanchez-Ruiz JM (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38**, 8138-8149
32. Pace CN, Laurents DV & Thomson J (1990) pH dependence of the urea and guanidine hydrochloride denaturation of RNase A and RNase T1. *Biochemistry* **29**, 2564-2572
33. Baxter NJ & Williamson MP (1997) Temperature dependence of 1H chemical shifts in proteins. *J Biomol NMR* **9**, 359-369
34. Zhang YJ, Caulfield T, Xu YF, Gendron TF, Hubbard J, Stetler C, Sasaguri H, Whitelaw EC, Cai S, Lee WC & Petrucelli L (2013) The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum Mol Genet* **22**, 3112-3122

35. Bentmann E, Neumann M, Tahirovic S, Rodde R, Dormann D & Haass C (2012) Requirements for stress granule recruitment of fused in sarcoma (FUS) TAR DNA binding protein of 43 kDa (TDP-43). *J Biol Chem* **287**, 23079-23094
36. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R & McKnight, SL (2012) Cell-free formation of RNA Granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767
37. Wang X, Zhou Y, Ren JJ, Hammers ND & Chapman MR (2010) Gatekepper residues in the major curlin subunit modulate bacterial amyloid fibril biogenesis. *Proc Natl Acad Sci. USA* **107**, 163-168
38. Shorter J & Lindquist S (2005) Prions as adaptive conduits of memory and inheritance. *Nature Rev Genetics* **6**, 435-450
39. White-Grindley E, Li L, Khan RM, Ren F, Saraf A, Florens L & Si K (2014) Contribution of Orb2A stability in regulated amyloid-like oligomerization of Drosophila Orb2. *PLoS Biol* **12**, e1001786
40. Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE & Wüthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *Pure Appl Chem* **70**, 117-142
41. Jung YS & Zweckstetter M (2004) Mars-robust automatic backbone assignment of protein. *J Biomol NMR* **30**, 11-23
42. Goddard TD & Kneller DG SPARKY. 3 Ed., San Francisco
43. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J & Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-698
44. Shen Y, Delaglio F, Cornilescu G & Bax A (2009) TALOS+, A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-233
45. Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378
46. Case D, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KJ, Onufriev A, Simmerling C, Wang B & Woods R (2005) The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668-1688
47. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJP, Chothia C & Murzin AC (2007) Data growth and its impact on the SCOP database: New Developments. *Nuc Acids Res* **36**, D419-D425
48. Hasegawa H & Holm L (2009) Advances and pitfalls of protein structural alignment. *Curr Opin Struct Biol* **19**, 341-348
49. Bai Y, Milne JS, Mayne L & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins Struc Func Gen* **17**, 75-86
50. Hvidt A & Nielsen SO (1966) Hydrogen exchange in proteins. *Adv Prot Chem* **21**, 287-386
51. Huyghues-Despointes BMP, Scholtz JM & Pace CN (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* **6**, 910-912
52. Myers JK, Pace CN & Scholtz JM (1995) Denaturant m-values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-2148
53. Li H, Robertson AD & Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704-721

54. García Mayoral MF, Martínez del Pozo A, Campos-Olivas R, Gavilanes JG, Santoro J, Rico M, Laurents DV & Bruix M (2006) pH-dependent conformational stability of the ribotoxin α-sarcin and four active site charge substitution variants. *Biochemistry* **45**, 13705-13718
55. Tanford C (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. in *Adv Protein Chem*, Academic Press, New York
56. Renner C, Schleicher M, Moroder L & Holak TA (2002) Practical aspects of the 2D 15N-[1H]-NOE experiment. *J Biomol NMR* **23**, 23– 33
57. Pawley NH, Wang C, Koide S & Nicholson LK (2001) An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of ^{15}N relaxation parameters. *J Biomol NMR* **20**, 149– 165
58. Tjandra N, Feller SE, Pastor RW & Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J Am Chem Soc* **117**, 12562– 12566
59. Bruschweiler R, Liao X & Wright PE (1995) Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886–889
60. Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. *J Am Chem Soc* **104**, 4546-4570
61. Cole R & Loria JP (2003) FAST-Modelfree: A program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* **26**, 203-213
62. Palmer AG, Rance M & Wright PE (1991) Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance carbon-13 heteronuclear NMR spectroscopy. *J Am Chem Soc* **113**, 4371-4380

5.8 Supporting Information

Sup. Figure 1. Evaluation of NTD expression and purification by gel electrophoresis.



This figure shows the levels of TDP-43 NTD expression in the total bacterial lysate (left lane) and following affinity purification on an Ni-NTA column (right lane). The samples were loaded on a 15% polyacrylamide gel and subjected to SDS-PAGE. The gel was stained with colloidal Coomassie blue. Molecular weight markers are shown in the middle lane.

Chapter 6

Ongoing work

This chapter provides an overview on the work that is currently being carried out regarding TDP-43 and amyloid energetics.

In previous chapters, structural results on both the N-terminal domain, and the C-terminal region of TDP-43 were presented. Using computational methods, we are working on integrating these results and data available from the literature into an atomistic molecular model for the full-length protein. This chapter explains the basic premises considered in building this model, the structural features which can be predicted, and a possible set of experiments to refine and validate the model.

One very surprising finding from the amyloid field is that highly polar sequences, such as GNNQQNY, can give rise to very stable amyloid structures. This presents a strong challenge to the dominant view held since Kauzmann's famous review [1] (see ref. [2] for a more recent update) that hydrophobic interactions are the main contributors to protein stability, since the amide groups that form hydrogen bonds within the protein will form H-bonds of similar stability with solvent water molecules, and therefore their net contributions will be small. Thus, deciphering the puzzle posed by the startling stability of polar amyloid will give novel and fundamental insights into protein conformational stability.

During the introduction of this thesis, the remarkable ability of diverse sequences to give rise to a similar amyloid fold was emphasized. Among the pathological amyloidogenic proteins, Aβ and Tau are mainly hydrophobic, polyQ and the TDP-43 C-terminal aggregative segment are highly polar and IAPP (linked to type II diabetes) is intermediate. Therefore, beyond its academic interest, comprehending the bases of amyloid stability which may well differ for polar versus hydrophobic amyloids, may well aid the future design of inhibitors. We are currently looking for differences between polar and hydrophobic sequences at the most fundamental level: the quantum nature of the interactions that provides cross-β spines' remarkable stability. Here, preliminary results on the basis of quantum mechanical calculations are shown with a brief description of the planned experimental setup for ongoing experiments.

6.1 Integrative model of full-length TDP-43

6.1.1 Introduction

Structural data on full-length TDP-43 are limited even though an increasing number of research groups focus on the study of this protein. In the past, a few laboratories have reported structural models of full length TDP-43 [3] but these were relatively unsophisticated and were reported before high-resolution structures became available for the N-terminal and RRM domains. Now that this information is available (**Fig. 1**), we expect that the structural model we aim to obtain here will provide a comprehensive perspective view which may aid our understanding of the interactions between the N-terminal and RRM domains and the C-terminal region, which have been recently been shown to be crucial for the native and pathological functions of TDP-43.

This chapter begins with a description of the procedures used to obtain and validate a computational model for this protein. All available data; namely, NOE-derived distances, dihedral, and H-bonds restraints for the solution NMR structures of the RRMs and N-terminal domains from the BMRB database are used. First, these data are used to perform simulated annealing in torsion angle space, as implemented in CYANA, to obtain a model of full-length TDP-43 using > 6000 distance restraints. Since this protocol does not impose any structural restraints on flexible regions for which experimental data are not available, afterwards set of Molecular Dynamics simulations are performed to obtain conformational ensembles for each of the independent, disordered segments.

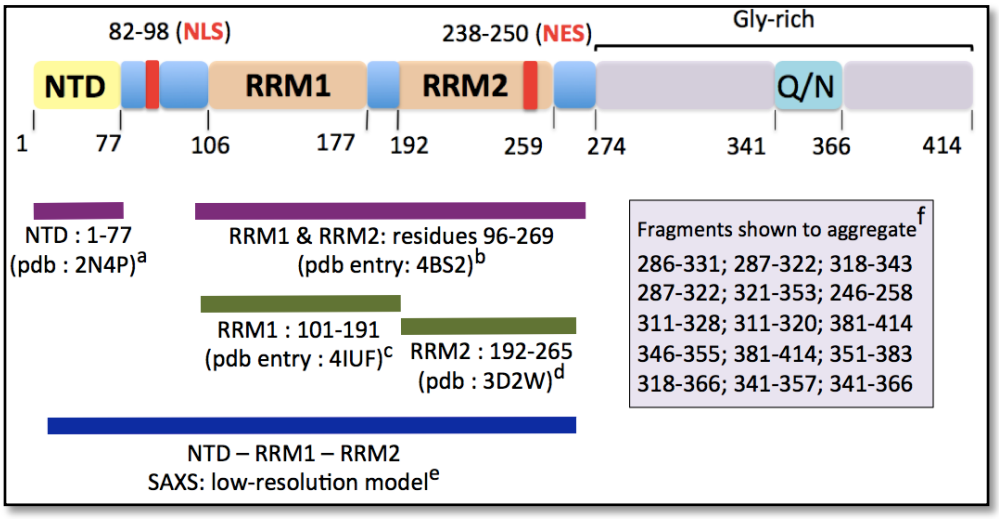


Figure 1. Structural information available for TDP-43 domains. NMR structures for the NTD, RRM1, and RRM2 domains are indicated with the purple bars, while X-ray structures are denoted with green bars. A low-resolution model for a homodimer consisting of the NTD and the two RRM domains was proposed on the basis of SAXS data (blue bar). To build the model for full-length TDP-43 we are only using NMR structures plus the structural model developed in Chapters 3 and 4 for the aggregative segment. The superscripts denote the corresponding references as follows: ***a*** for ref. [42]; ***b*** for ref. [39]; ***c*** for ref. [41]; ***d*** for ref. [40]; ***e*** for ref. [14], and references for ***f,*** as well as the main findings with those fragments, where already discussed in Table S1 (Chapter 4, Section 4.8.3).

Finally, twelve tandem repeats of the 341-366 segment are attached to the C-terminus of the protein to mimic a cellular model that is described in the literature [4] to reproduce the hallmarks of ALS in experiments in cells. These additional aggregative segments are modeled based on the structural model developed in Chapter 4 for residues 341-357; in the context of the 341-366, I found that the 358-366 residue segment acts as a linker to connect successive β-hairpins. Interestingly the length of this linker suffices to connect two β-sheets into a cross-β spine structure at their experimental inter-sheet distance that we determined by X-ray diffraction. In addition, since the dimerization of the N-terminal domain has been reported to be important for TDP-43 physiological functions and pathological aggregation [3], I consider possible modes of dimerization of the N-terminal domain. The effects of mutations aimed at this disrupting this domain's stability and their implications for TDP-43 normal function and pathogenesis will be studied in the near future. In this regard, possible strategies for validation and refinement are discussed.

6.1.2 Preliminary results

Full-length TDP-43 computational model: fragment-assembly approach – The first step towards a full -length model of TDP-43. The high-resolution NMR structures for the N-terminal and the two RRM domains of TDP-43 are deposited in the Protein Data Bank (PDB) under accession codes 2N4P and 4BS2, respectively. Although it is possible to simply merge both pdb files, to facilitate later steps and as an additional confirmation of the published RRM domains, we used the deposited chemical shift data (BMRB entry: 19290) to independently calculate dihedral angle restraints using the TALOS+ program [5], as they were not found in the original publication. Once this set of restraints is generated for the RRMs, all the structural information was ready to be submitted along with the NOE-derived and H-bonds restraints to the CYANA program [6] for a standard simulated annealing structure calculation. The first structural model that fully agrees with the input data (*i.e.* it has no constraint violations) is shown in Figure 2. The Q/N segment 341-357 was included as a β-hairpin on the basis of our results in chapters 3 and 4.

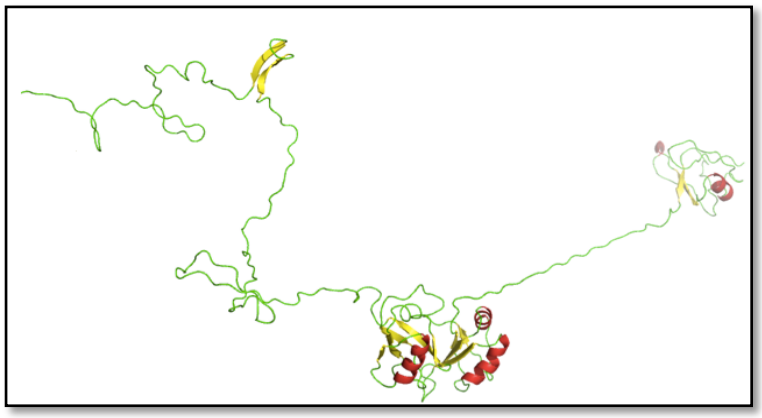


Figure 2. Preliminary model for the backbone of full-length TDP-43. The first model obtained for the full-length protein only includes backbone restraints for the NTD and the RRMs domains, and residues 341-357 are restricted to form a β-hairpin. α-helices and β-strands are shown in red, and yellow, respectively.

This first approach confirms that the experimental data for these structural domains can be successfully used to build a model for the full-length TDP-43, although the rest of the protein is chiefly unfolded as a consequence of the lack of experimental restraints. Since these other segments are intrinsically disordered, but modeled as artificially extended segments by CYANA [6], we address a plausible way for generating more realistic disordered conformational ensembles in the next sections.

Structural model for the TDP-43-12xQ/N pathology-mimicking construct - Previous results from the Buratti/Baralle laboratory have shown that TDP-43-12X-Q/N, which contains a short N-terminal "Flag" tag for antibody recognition and a dozen Q/N repeats at the C-terminus, is an excellent construct for recapitulating and studying pathological TDP-43 in cells since: i) it is able to recruit endogenous protein into aggregates; ii) it induces loss-of-function effects, and iii) the trapped aggregates undergo the post-translational modifications which are hallmarks of the pathology. Therefore, we have attached twelve repetitions of the 341-366 segment to the C-terminus of the protein. To build this model, we used the same approach as for full-length TDP-43 (*vide supra*), except that now the repetitions are included with the restraints derived from the amyloid-like conformer described in Chapter 4. We built two β-sheets, each composed of six β-hairpins, resulting in 24 β-strands that form a fibrillar fold. Fig. 3 shows the result of the initial calculation. Interestingly, residues 358-366 act as a linker with the correct length to connect 341-357 β-hairpins allowing proper cross-β spine formation, as illustrated in **Fig. 4.** This linker serves to both connect the six β-hairpins of each sheet, as well as to allow the two sheets to pack against each other preserving the integrity of the fold proposed in Chapter 4 for this prion-like region.

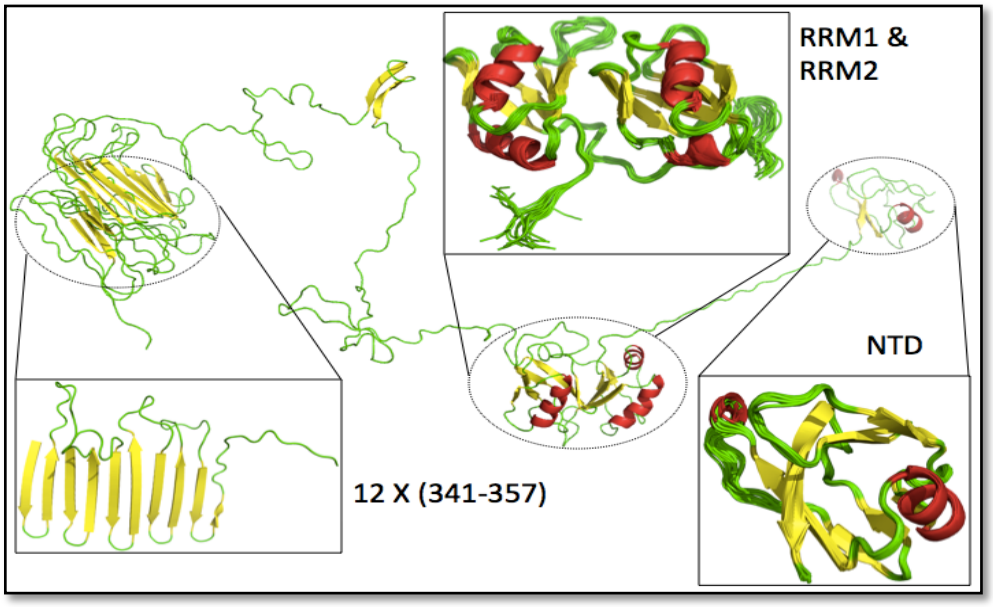


Figure 3. **Preliminary model of full-length TDP-43-12xQ/N.** The model shown in **Fig. 1** is now extended by adding the twelve Q/N tandem repetitions attached to the C-terminus of the protein with the proposed parallel β-turn topology. The structured domains are magnified, showing superposition of 20 conformers for the NTD and RRMs domains. One of the two mating sheets formed by six 341-366 repetitions is displayed to illustrate how the 358-366 segment effectively connects the β-hairpins.

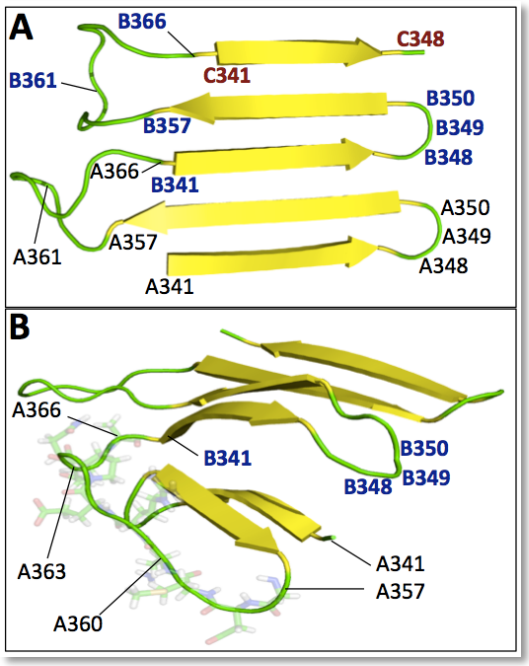


Figure 4. Residues 358-366 enable oligomerization into a cross-β spine structure. We showed in Chapter 3 that the segment 341-357 could adopt a β-hairpin conformation which is followed by residues 358-366 which assumes an ordered, noncanonical structure. This segment not only allows β-sheet growth with the parallel β-turn configuration (**A**), but also provides a way to symmetrically fold the twelve repetitions, so that a cross-β spine in which each mating sheet contains six monomers is formed (**B**). When connecting the two β-sheets, residues 358-366 retain their antiparallel configuration with respect to the β-hairpin.

MD simulations are used to generate disordered conformational ensembles for the unstructured regions – The lack of restraints in the C-terminal regions and segment connecting the NTD and RRM1 results in long, extended chains for these segments. Although there is evidence for them to be intrinsically disordered, these regions will be much more compact because their backbones will adopt conformational ensembles which include other regions of the Ramachandran plot in addition to extended conformers and due to molecular crowding effects [7]. Moreover, residues 82-98 constitute the Nuclear Localization Sequence (NLS), and it was reported that the corresponding NLS in FUS can adopts an α-helical conformation [8, 9] (especially when bound to nuclear import factors). Given the similar functions of TDP-43 and FUS, and considering that they both are linked to ALS, we searched in the literature and found NMR evidence for the NLS sequence in TDP-43 to be chiefly unfolded [10, 11]. Accordingly, we did not impose specific structural restraints for this segment.

The NES of TDP-43 is located within the RRM2, and contains an α-helix and a β-strand. Since the NES from FUS/TUS, though not other hnRNPs, was reported to populate α-helical conformations [12], we wondered whether these observations in TDP-43 might be a consequence of nucleic acid binding, as the restraints that we use for this region corresponds to an RRM1-RRM2 construct in complex with RNA. Searching in the literature, the only additional structure for this RRM2 containing the NES, is an X-ray crystal structure which is bound DNA and displays a slightly longer α-helix (**Fig 5**). Considering that this structure corresponds to mouse TDP-43 rather than to the human protein, and their slightly different sequences in this region, yet conserves the structural elements strongly suggests that the α-helix and the β-sheets are characteristic features of this segment (**Fig. 5**). In fact, we found NMR structures deposited in the protein data bank for the independent RRM domains in the absence of bound nucleic acid that support this idea. Note that they were not included in the previous Fig. 1, as we could not find literature references for them although they are deposited in the PDB. Nonetheless, they seem to be well-determined on the basis of their available NMR restraints, and we compare the NES conformation in both the isolated and nucleic acid-bound structures, along with the mouse isoform, and they are all in good agreement. Therefore, we did not impose additional restraints to this region for modelling the structure of full length TDP-43.

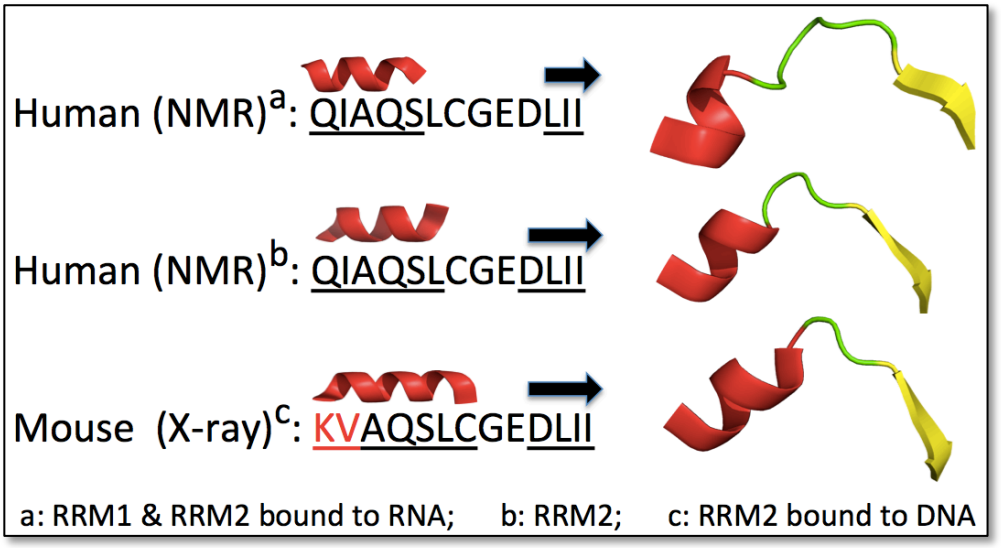


Figure 5. Secondary structures adopted by the NES in TDP-43. The Nuclear Export Signals (NES) of TDP-43 contains short α-helical and β-sheet segments. Comparison with RNA-bound [39] and unbound (pdb: 1WF0) NMR structures show that these structural features are not a consequence of nucleic acid binding. We compare these regions with the NES of mouse TDP-43 RRM2 bound to DNA[40] and found the same structural arrangement, despite variations in their sequences. Note that the two non-conserved residues in mouse are marked in red.

With all these considerations, we are currently performing MD simulations to determine reasonable conformational ensembles for the following segments of the structure: i) Pro78-Ala99, which contains the NLS and connects the NTD and the RRM1; ii) Ser226-Ser317, which connects the RRM2 and the hydrophobic patch in the C-terminal region, and iii) the segment Phe367-Met414 that corresponds to the last residues in the C-terminal domain.

These simulations are intended to downsize these extended segments in a plausible way (Fig. 6), since they all will otherwise be in extended β-strand-like conformations that have artificially low conformational entropy. MD allows the chain to adopt a more realistic ensemble of conformations with a less artificial combination of phi/psi angles, as shown in Fig. 7. This results in higher conformational entropy and lower free energy, while avoiding steric clashes. Whereas it is becoming possible to fold short protein segments which natively adopt well-defined folded states by MD [13], very long and more sophisticated calculations would be needed to attempt to fold the chains into their native intrinsically disordered conformational ensembles, and this is beyond the scope of our current investigations.

It is important to note that the linker connecting both RRM domains (residues 180-190) was not observed in the free protein, but is resolved when these domains are bound to RNA. This indicates that they adopt a more rigid conformation upon nucleic acid binding, nevertheless, we have not taken this into consideration in our first unrefined , preliminary models.

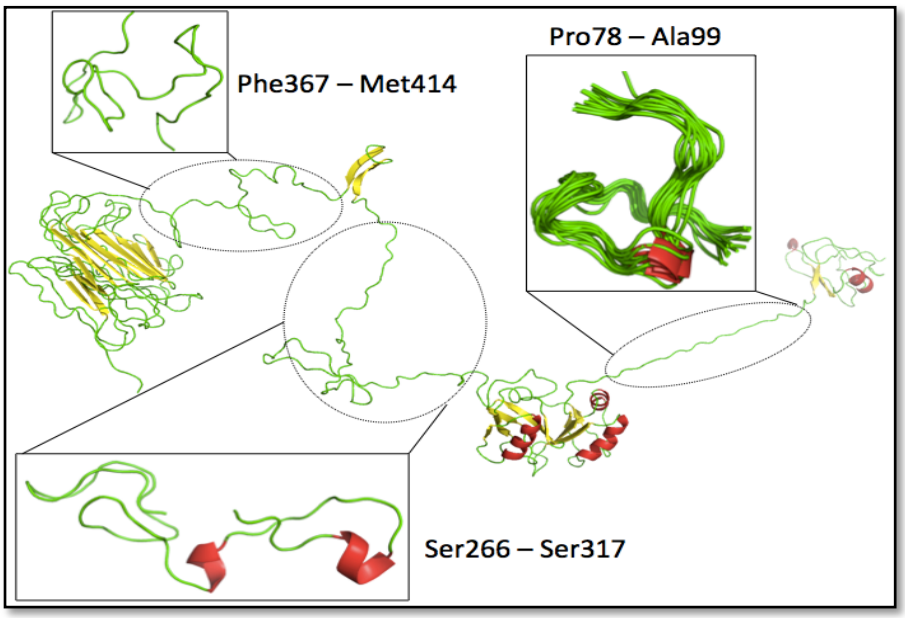


Figure 6. MD simulations to condense extended chains. After applying molecular dynamics, the regions of extended chains (*within the ovals*) adopt ensembles of conformers that are generally more compact (*boxed zoomed views*). Residues 78-99 now adopt a more compact structural ensemble with some turn-like features. This segment corresponds to the NES, and literature data suggests that it is disordered but may adopt a helical conformation upon binding to nuclear export factors. The same procedure is being applied to the C-terminal segments 266-317 and 367-414, flanking the Q/N-rich region 341-357, which adopt more compact ensembles with short helix-like features in the case of the 266-317 segment. All the simulations are being performed with the exactly same set of parameters used for MD as described in Chapters 2, 3, and 4.

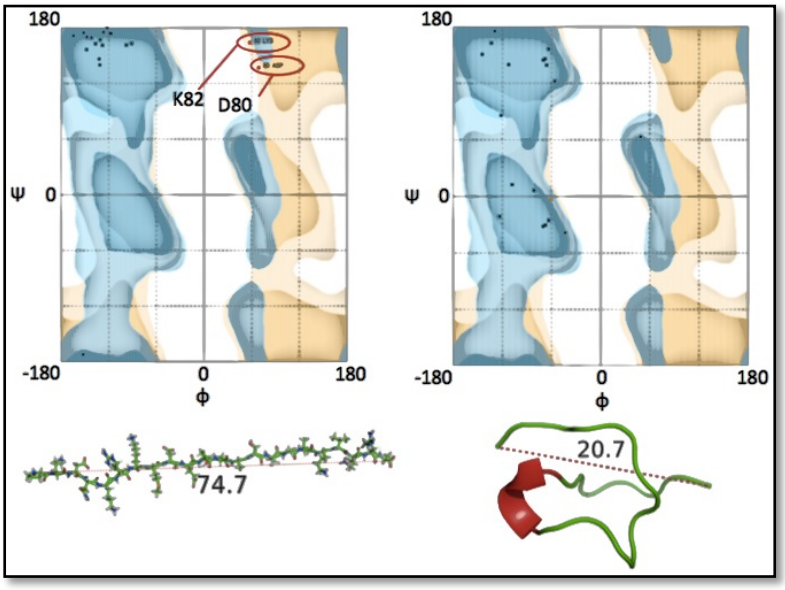


Figure 7. Ramachandran map of the extended and condensed Pro78 – Ala99 segment. The standard annealing using experimental restraints for folded domains of TDP-43 does not impose any structural preference for disordered segments that lack experimental data. The phi/psi angles of these regions correspond to large linear chains (*left*). Note that two residues (K82 and D80) lie in disfavored regions of the Ramachandran map. After applying molecular dynamics, these segments adopt more compact ensembles with all the residues in allowed regions (*right*).

TDP-43 forms dimers through the NTD – It has been shown that the N-terminal domain of TDP-43 is essential for the protein's dimerization. However, information regarding the dimeric structure is scarce. In principle, this assembly could be mediated by disulfide bond formation as each NTD contains two surface exposed Cys residues, or by non-covalent contacts. Since we do not have structural insights for this process yet, two possible binding modes have been considered in a scenario where dimer formation is mediated by the oxidation and formation of disulfide bonds between the domains' exposed cysteine residues, as shown in **Figures 8A** and **8B**.

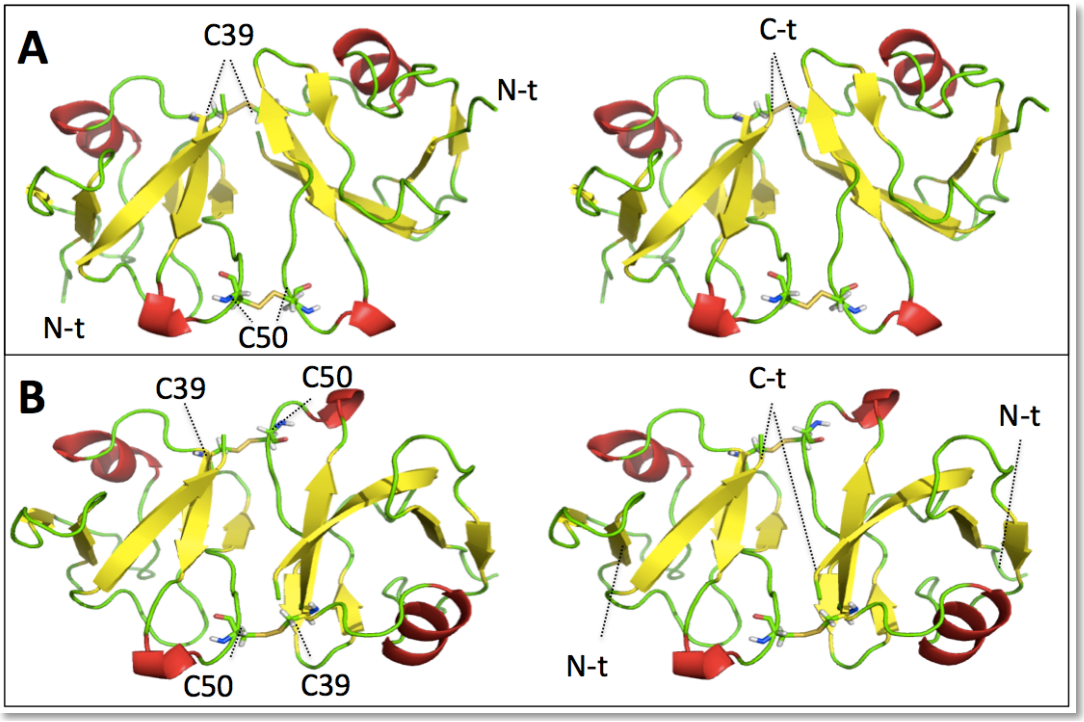


Figure 8. NTD dimerization through disulfide bond formation. (A) This first model of dimerization consists of symmetric disulfide bond formation, where the same exposed cysteine residues are oxidized to for the dimer. **(B)** Here, dimerization is considered to be antisymmetric between the different cysteine residues of each NTD.

We are running MD simulations to study whether one of these possibilities display more favorable van der Waals or charge-charge interactions, although we are attempting to obtain experimental evidence for the correct dimeric structure.

By SAXS, a low-resolution envelope of the TDP-43 homodimer has been determined, which permits the localization of the relative position of the domains [14]. In this model, the two NTDs would contact each other and the loop connecting beta strands 2 and 3 of the RRM1 domains would be rather close together. The RRM2 domains and the C-terminal domains of the two subunits are relatively far away from each other. This is illustrated in **Fig. 6A**, and we are using the shape of that SAXS envelop to refine our model, as shown in **Fig. 6B**. The elucidation of the NTD-NTD dimerization mechanism along with the fitting to the SAXS data will likely result in a more refined model.

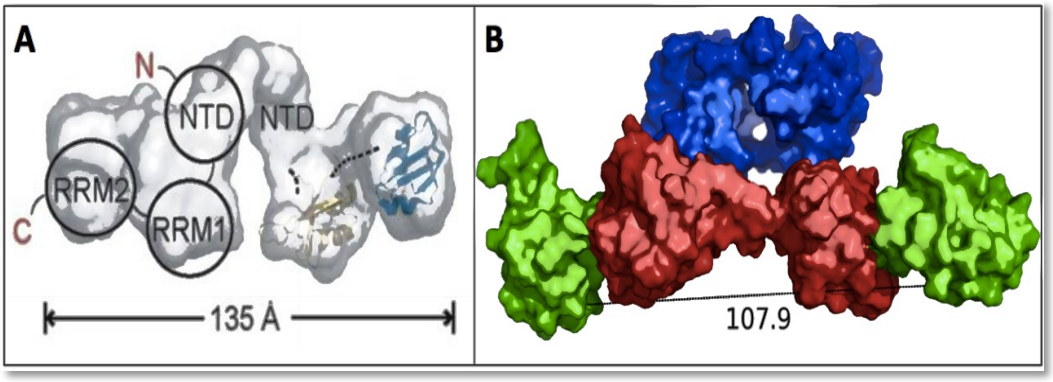


Figure 9. Model for dimerization based on SAXS data. A model for a homodimer consisting of the NTD (blue), RRM1 (red) and RRM2 (green) domains is being refined on the basis of SAXS data. The model shown here is built based on the antisymmetric disulfide bond formation shown in **Fig. 8B,** just to illustrate the procedure that we pursue. Structural data regarding NTD-NTD homodimer formation will allow a proper modeling of the system, and likely a better fit to the SAXS envelope (*Panel A* is adapted from reference [14]).

Possible dimerization of TDP-43 & TDP-43-12xQ/N – On the basis that TDP-43 exists as homodimers, we are considering in our models a plausible assembly of wild type TDP-43 with the construct that contains the twelve Q/N repetitions attached to the C-terminus of the protein. This dimeric model is being built with the above considerations for homodimer formation, where the only difference is the dozen Q/N repeats that drive oligomerization into an amyloid-like nucleus.

Recruitment of endogenous TDP-43 into aggregates within this cellular model can be tentatively attributed to two different processes. First, the tandem repetitions can form an amyloid fold that might act as a sponge and sequester nuclear TDP-43 through its Q/N segment. Secondly heterodimerization *via* the NTDs of endogenous TDP-43 with TDP-43-12XQ/N would increase the local concentration of the Q/N rich segments and thereby favor their self-association. Currently, we do not know if both processes act independently or synergistically, and we expect that this model could explain some essential interactions to help in understanding these mechanisms. **Figure 8** shows a first calculation in which the NTD mediates dimer formation through disulfide bond formation, whereas **Figure 9** shows what this assembly could look like on the basis of SAXS. Again, the true mode of dimerization needs to be elucidated to achieve a more refined model, but these figures give an overview on the current state of the work.

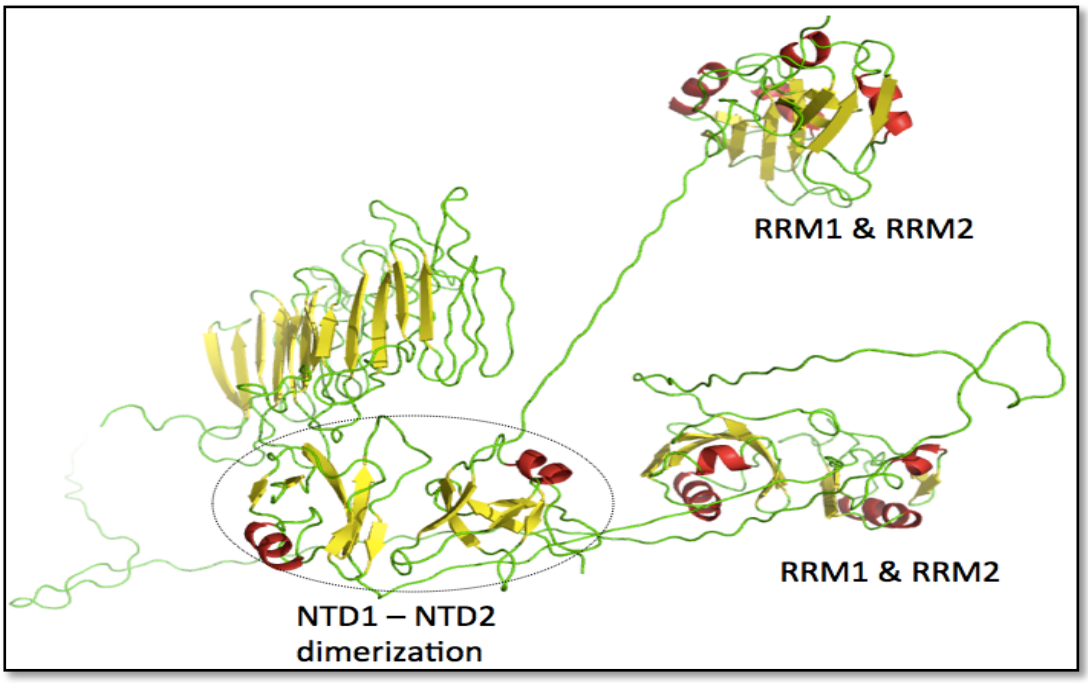


Figure 10. Model for a TDP-43 — TDP43-12xQ/N dimer through their NTDs. Models from **Figures 2 and 3** are now merged into a dimer mediated by assuming heterogeneous disulfide bond formation between the respective NTDs (Figure 8B). How dimerization really occurs remains to be determined and this model has not yet been subjected to MD to produce more compact and realistic conformational ensembles for the disordered segments.

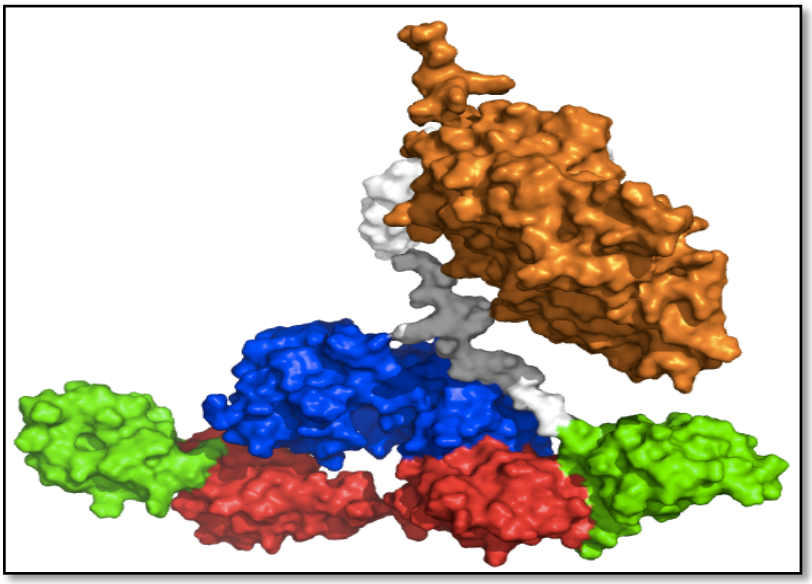


Figure 11. Model for a TDP-43 — TDP43-12xQ/N dimer through their NTDs. The plausible dimer between wild type TDP-43 and the construct with the Q/N repeats shown can be refined with the SAXS envelope. Here, the NTD (blue), the RRM1 (red), and the RRM2 (green) are shown along with the 12xQ/N repeats (orange). The C-terminal region has been truncated for display purposes (colored in white). In this figure, one can appreciate that twelve tandem repetitions of the Q/N-rich region folded in the amyloid-like conformation occupy a large region of space, suggesting that it could readily act as a sponge to recruit endogenous TDP-43 into these aggregates.

6.1.3 Discussion and ongoing work

The goal of this modeling effort is to develop a structural framework for the complete TDP-43 proteins. We have emphasized that the mechanism for dimer formation mediated by the N-terminal domain needs to be addressed, and we are working in this direction. In parallel, a deep examination of the C-terminal domain properties is still necessary to elucidate the structure of the putative functional amyloid, and its possible conversion into pathological amyloid. These results can be incorporated into our model as they become available. In the future, this will advance our understanding of the origin of the loss-of-function induced by the 12xQ/N repetitions and the role of the N-terminal domain in mediating dimerization, and WT TDP-43 /TDP-43(12xQ/N) interaction and recruitment into aggregates.

The ^{1}H, ^{15}N & ^{13}C chemical shifts and the high-resolution structure of the N-terminal domain open new possibilities to understand TDP-43-mediated processes. This structure is stable *in vitro,* however, high resolution data on the conformation and stability TDP-43 in cells are much more scarce although it is known that the protein is continuously produced, that its concentration is tightly controlled, and that the protein is rapidly turned over in the context of Drosophila cells [15]. To fill this gap, our collaborators in Trieste (EB & FB) are designing a series of mutants with the aim of disrupting this domain's stability. Analyses of the behavior of these TDP-43 variants in cells and of their structure and stability in our lab using NMR should allow us to test whether or not that the structure and stability of the NTD within cells and *in vitro* are similar.

6.2 A ubiquitous hydrogen-bonding feature in amyloids: unraveling the kernel for aggregation

6.2.1 Introduction

As noted previously, most experts believe that hydrophobic interactions make the chief contribution to protein stability since they assume that protein/protein H-bonds are similar in stability to protein/solvent H-bonds [2]. Nevertheless, the study of protein variants carrying mutations designed to disrupt H-bonds within proteins support the idea that these H-bonds can substantially stabilize protein structures [16, 17].

Starting in the 1990's, interest in hydrogen bonds from a computational perspective has been rekindled and additional classes of H-bonds, sometimes involving unusual atoms like carbon, or with peculiar energetic properties, such as resonance-assisted hydrogen bonds (RAHBs) [18, 19] or low barrier hydrogen bonds (LBHBs) [20, 21] have been proposed.

One of the most interesting properties is the mutual strengthening experienced by a hydrogen bond in the presence of similar interactions. This is the so-called hydrogen bond cooperativity (HBC), a feature that has been widely observed. This property is of special importance in peptides and proteins, where the contribution of Dannenberg's group havs been of special importance [22–24], as they pioneered the extension the HBC concept to pathological amyloidogenic sequences [25, 26]. Previous works indicate that as nascent amyloids grow, the strength of the H-bonds linking the upcoming β-strand to the growing oligomer increases with the number of strands already within the nascent amyloid, up to a certain number where a maximum is reached. For instance, Baker *et al*. studied this effect in the amyloid-forming segment GNNQQNY of the yeast prion protein Sup35. They found that up to four monomers, each of the strands was more strongly hydrogen-bonded to the growing β-sheet, and they attributed this effect to a hyper-polarization of the interactions [27]. They showed that pure DFT calculations could be applied on these systems by using periodic boundary conditions.

It is intriguing that amyloid aggregation occurs in segments whose amino acid residues have widely different chemical properties: from predominantly hydrophobic sequences, such as the Aβ peptide involved in Alzheimer Disease to highly polar segments, as in the case of polyglutamine, TDP-43(341-357) or the heptapeptide GNNQQNY from the yeast prion protein Sup35. Moreover, the configuration of the β-strands in the aggregates have been found to be either parallel or antiparallel, with at least eight different possible amyloid β-sheet topologies [28]. This prompted us to search for common patterns that might be shared between sequences of different nature (polar *vs*. hydrophobic) and configuration (parallel *vs*. antiparallel) towards finding a common feature that can drive aggregation to yield the same quaternary fold.

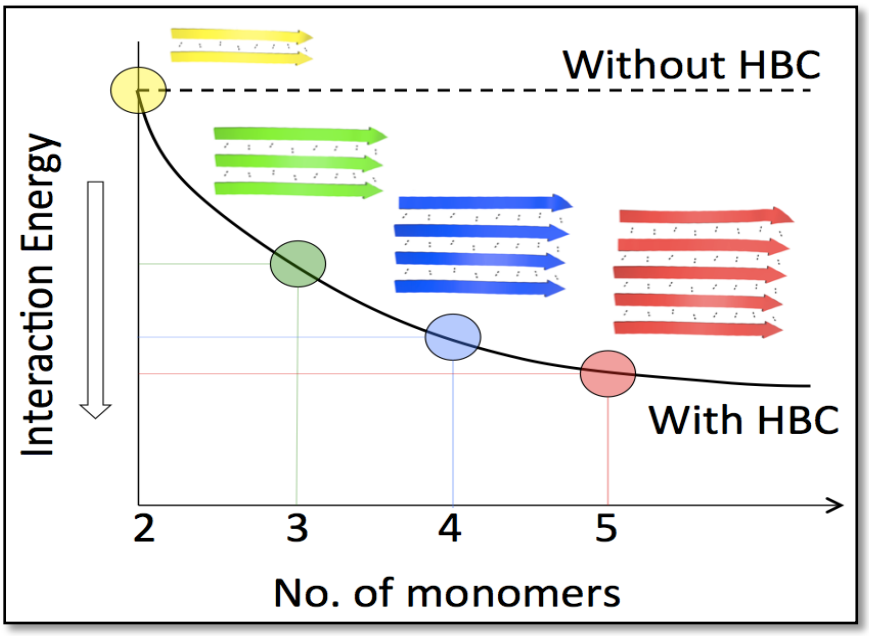


Figure 12. Graph illustrating the HBC effect in amyloids. Cooperative H-bonds occur when the incoming strands dock the growing β-sheets with stronger interaction than that corresponding to isolated strands. In this figure, two isolated strands (yellow) interact with a value. Since all the monomers are the same, a third unit would interact with the two-stranded β-sheet with the same energy value. However, in this situation (green sheet), the third monomer forms more stable hydrogen bonds. Binding of additional β-strands is increasingly favorable and reaches a limiting value for very long sheets.

Previous studies lack a proper description of cooperativity – The vast majority of these studies have focused on the interaction energy of an edge β-strand hydrogen-bonded to an N-stranded β-sheet, corresponding to the formation of the (N+1)-stranded β-sheet. In this framework, with increasing values of N, it is possible to study the strengthening of the hydrogen bonds upon oligomerization (*i.e.*, the HBC). However, much less is known to what happens to the HBs formed by the β-strands in the middle of the amyloid as this process occurs. Kobko *et al*. considered this situation in a rigorous study of formamide chains [22], in an attempt to approach the HBC in proteins. They found that not only incoming formamide molecules form stronger hydrogen bonds (variation with oligomer size), but also that those already formed were strengthen (variation with position). These results are shown in Fig. 13 and although they were interesting, the molecules studied and geometry imposed that they imposed does not resemble that of a restrained peptide backbone. Nevertheless, they related this HBC to the cooperativity observed for polyalanine [29] and polyglycine [30] α-helical peptides.

Rosetti *et al*. studied the HBC in poly-Glutamine β-sheets, investigating the effect of the intermediate strands in these fibrils to account for the HB position [31]. They used semi-empirical methods, and showed that these computational approaches, which are less time-demanding than DFT, can be used to study this phenomenon. However, they limited their analysis to HB distances. Although the use of these geometrical criteria is a common procedure in studying HBC, there is no clear relationship between the shortness of HBs and their strength, since there are pairs of HBs that have the same energies but different O···H distances [22, 32].

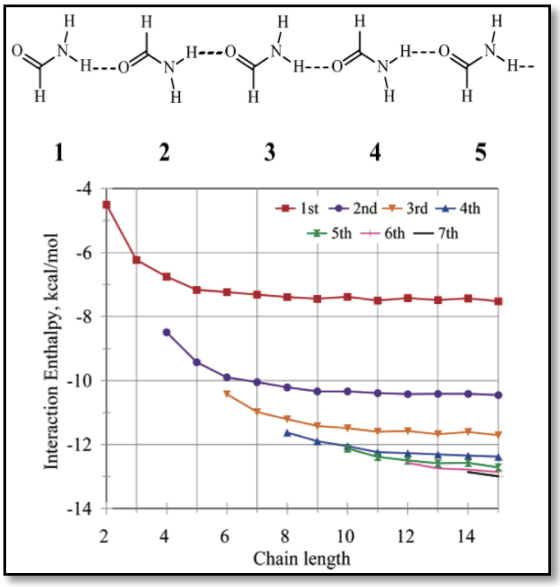


Figure 13. HBC in formamide chains. Interaction energies computed with respect to the chain length revealed that the incoming formamide chain forms stronger hydrogen bonds. The plot shows seven curves (data sets) corresponding to the first seven HBs formed within up to 15 formamide molecules linearly arranged. The first data point of each set correspond to a molecule docking on a system composed of N formamides, showing the cooperative effect (HBC) described in the text. Interestingly, within each data set, the interaction of the same HB is evaluated for larger formamide chains and HBC is also observed to a lesser extent. This figure is reproduced from [22].

Established concepts to be revisited – A plausible explanation for the HBC in amyloids' β-sheets was attributed to the presence of Gln & Asn side chains that provide the assembly with additional amide groups able to reinforce the hydrogen-bonding network. Dannenberg's group studied different systems consistent of β-sheets and found that segments containing these polar residues exhibited an enhanced HBC. They ascribed this observation to a distortion of the inter-strand backbone-backbone HBs to afford a more optimal arrangement of the sidechain – sidechain H-bonds from Asn and Gln residues. This prominent role of amide-containing sidechains is supported by the observation of poor or negligible HBC in non-polar sequences that lack these residues, either by *in-silico* mutations to Ala on model systems, or by studying segments from predominantly hydrophobic amyloids [25, 33]. A general observation is that β-sheets twist upon geometry relaxation regardless of their residue compostion and according to Danneberg's results, polyalanine and polyglycine β-sheets do not display HBC. This reinforces the idea of Gln and Asn are responsible for compensating backbone distortions and suggests that they are directly involved in HBC. However, there are no experimental structures for those model systems to infer their twisted nature, and we showed in Chapter 2 that GNNQQNY displays HBC in both the experimental X-ray structure and in the twisted conformation resulting from geometry optimization, which strongly suggest that this phenomenon is not a compensation effect arising from sidechain optimal hydrogen bonds. Along with these drawbacks, most of the previous studies also ignored solvation effects, and dealt with gas-phase systems. Whereas this is convincing in terms of studying the intrinsic reactivity of amyloid's HBs, the polarization of the solvent should be included since the nature of the HBC effect is purely electronic.

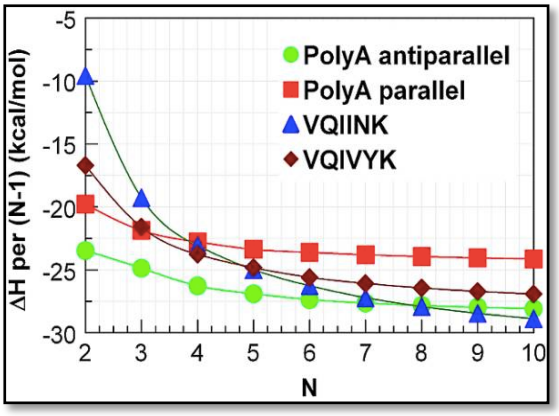


Figure 14. Gln and Asn side chains are proposed to contribute to HBC. β-Sheets composed of polyalanine display poor HBC, as the energy difference between the first two values in these systems is slightly over 1 kcal/mol, which is the assumed error for these calculations. However, when the β-sheets contain Q and N residues, a noticeable HBC is observed, suggesting a special role of these residues (see main text). This figure is reproduced from [26].

6.2.2 Preliminary results

Interaction Energies from amyloid structures in the PDB: model setup – To study cooperative effects in amyloid formation, we begin utilizing a model in which hydrogen bonds are analyzed from Kobko *et al*'s perspective applied to formamide chains. Whereas in that case only one hydrogen bond was present between each pair of formamide molecules, here two β-strands encompass a number of these interactions, and a direct estimation of the HB energy between them can be obtained from their interaction energy, which we denote here as *Hydrogen Bonding Strength (HBS)*. To study the origin of the cooperative effects, we have chosen two completely different amyloid structures: a highly polar, Asn- and Gln-rich sequence arranged in parallel β-sheets, and a predominantly hydrophobic segment that has been shown to form antiparallel β-sheets. The former corresponds to the heptapeptide sequence GNNQQNY from the yeast prion protein (pdb: 1YJP) [34], whereas the latter is a segment from the Aβ peptide involved in Alzheimer Disease (Aβ35-40) (pdb: 2ONA) [28], whose sequence is MVGGVV. The corresponding β-sheet structures span up to five β-strands, designed as ***ABCDE***, where each capital letter corresponds to each of the mutually hydrogen-bonded β-strands. Therefore, ***AB*** would be a dimer, ***ABC*** a trimer, and ***ABCD*** a tetramer. The simplest case is that of a two-stranded β-sheet (***AB)***, as only one HBS value can be computed, corresponding to the interaction between strands ***A*** & ***B***. Within this approach, if one considers a trimer ***ABC***, it is possible to select two hydrogen-bonded pairs of β-strands to compute their corresponding HBS values: ***AB,*** & ***BC***. In the case of a four-stranded β-sheet ***ABCD***, there are three possibilities corresponding to selection of the ***AB, BC,*** & ***CD*** β-strand pairs. Finally, for the pentameric system ***ABCED***, four values for the HBS could be obtained: ***AB, BC, CD*** & ***DE.*** This means that the result of docking additional β-strands to the initial ***AB*** dimer allows the

computation of N-1 interaction energy values, being N the total number of β-strands in the oligomer (from 2 to 5).

To maintain the parallelism with Kobko's formalism, we studied how HBS values are altered in the framework of a growing oligomer. The simplest case would correspond to two (AB) and three (ABC) formamides, where one compares the increase in AB···C strength with respect to that of A···B. This relates to cooperative effects, and the next obvious calculation would involve obtaining the energy of the ABC···D hydrogen bond. The interpretation is straightforward on the basis of **Figs. 12 and 13** (***vide supra***): this value increases non-linearly as the chain length is longer. The novel approach of Kobko was to study how HBs were affected as a function of their position, rather than only considering the incoming unit at the edge. This corresponds to comparsion of A···B with respect to A···BC, A···BCD, and so on, for each H-bond at any position. In our approach, each β-strand corresponds to each of those formamide molecules and, in contrast to that situation where only one HB hold two molecules together, here we replace that single interaction for the HBS term. With this framework in mind, we proceeded and studied the interaction of several systems as illustrated in **Fig. 15**.

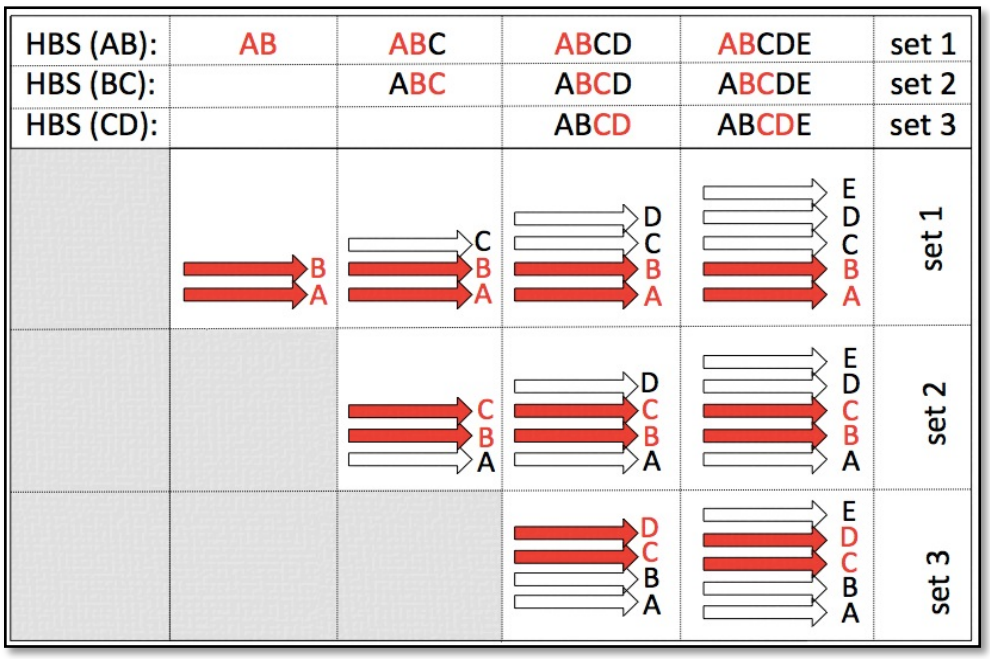


Figure 15. Set up for studying the HBC effect in amyloids. To analyze the impact of β-sheet elongation on the hydrogen bonds between the β-strands, we computed the interaction energies for selected pair of monomers. We call this energy value *Hydrogen Bonding Strength* (HBS) and we use it to study possible cooperative effects in experimentally resolved amyloid structures. Regardless of the concrete system, the analysis is performed at two levels; namely, i) the strength of the interaction of incoming strands onto a preformed sheet, and ii) the effect on a preformed network of HBs within a selected pair of β-strands.

HBC encompasses two different contributions – According to Fig. 15, three sets of values have been obtained in our calculations. Set 1 consists of the ***A···B*** HBS in a situation where the dimer is isolated (AB), and when one, two, and three additional strands dock onto it (that is, ***A···B***C, ***A···B***CD, and ***A···B***CDE being C, D and E the incoming strands). Similarly, set 2 refers to the A***B···C***, A***B···C***D, and A***B···C***DE HBS values, whereas in set 3 AB***C···D*** and AB***C···D***E were studied. Note that strands in **bold** and *italics* correspond to those shown in red color in **Fig. 15**.

Comparison of sets 1, 2, and 3 for the GNNQQNY sequence (Fig. 16 left panel) shows that the expected HBC gets more intense from set 1 to 2, and from set 2 to 3. This means that the HBS is higher as the oligomer grows, as expected according to the traditional description of the HBC effect. The right panel in Fig. 16 shows the same trend for the MVGGVV sequence, indicating that HBC might not be restricted to Gln- and Asn-rich sequences. This result provides new insights, since previous studies concluded that hydrophobic segments do not show cooperative H-bonds [26]. Another interesting observed feature is that only in the case of GNNQQNY, cooperative effects occurs within each set (black, red, and green curves in Fig. 16). This was already observed in formamide chains, as described above. The interpretation is that a given set of HBs at a fixed position within a selected pair of strands is strengthen when additional monomers elongate the fibril. That is, the A···B interaction strength increases as A···BC < A···BCD < A···BCDE. Likewise, the same occurs with B···C in AB···C < AB···CD, < AB···CDE or for C···D in ABC···D < ABC···DE. However, this effect is not observed for the non-polar sequence MVGGVV, where no significant variations within the same set of values occur as in the case of GNNQQNY. We hypothesize that there are two types of cooperativism: a ubiquitous HBC, which we called *residual HBC* (re-HBC) that is common to all amyloids, and an additional *hyper HBC* (hy-HBC) that might be a singular feature of Q/N-rich sequences.

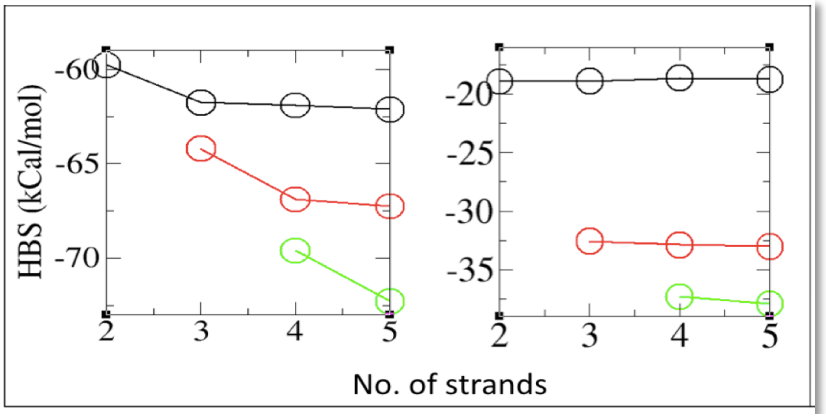


Figure 16. Two types of HBC in amyloids. (Left) Interaction energy values related to the Hydrogen Bonding Strength (HBS) that hold each pair of GNNQQNY strands together. HBC occurs as in the case of formamide chains. **(Right)** The same magnitude computed for MVGGVV shows that this peptide's HBC is similar to GNNQQNY's, which means that this effect is not restricted to polar sequences. For this reason, we termed it *residual HBC*. The analysis of sets 1, 2, and 3 from **Fig. 15** (black, red, and green curves, respectively) reveals that MVGGVV do not show the intra-set cooperativity seen in GNNQQNY, and we call this contribution *hyper HBC*.

HBC does not depend on configuration, sequence or length – Considering the substantial differences between the two systems for which the calculations were performed, the presence of a *hy-HBC* effect could be attributed to three different factors; namely, the parallel or antiparallel configuration, the hydrophobic/polar nature of the peptides and the *effective number of hydrogen bonds*. Whereas both are composed by roughly the same number of residues, GNNQQNY is able to form five additional H-bonds due to the amide groups of its Asn and Gln side chains. Therefore, it seems plausible that the effective number of hydrogen bonds (11 for GNNQQNY *vs*. 5 for MVGGVV) is the main factor responsible of these differences. Another possibility could be the different orientations (parallel vs. antiparallel) of the β-strands in the β-sheets, although this seems contradictory since H-bonds in antiparallel β-strands have a more optimal geometry since we are using the experimental structures to avoid twisted models. To investigate whether this is indeed an effect of Asn/Gln sidechains or could be rationalized in terms of the number of HBs or the strand configurations, we performed the same calculations on a hydrophobic segment with sequence LVFFAEDVGSNKGAIIGLMVGGVVIA that corresponds to residues 17-42 of the Aβ peptide involved in Alzheimer's disease (Aβ17-42, pdb code 2BEG) [35], keeping the discussion on the basis of structures that have been validated through deposition in the PDB. This segment is an excellent candidate to study the origin of *hy-HBC*, as it is a longer (25 residues) hydrophobic amyloid-forming peptide with a similar number of H-bonds as Sup35. Therefore, it should be likely to behave as the polar GNNQQNY amyloid, if cooperativity is related of the *effective HB number* that is to be formed. This means that if *hy-HBC* was not detected in MVGGVV because of its low number of inter-strand hydrogen bonds (only 5 versus 11 in GNNQQNY, where the effect is observed), the HBS profiles of Aβ17-42 should show this hy-HBS behavior. **Fig. 17** shows the effective number of inter-strand HBs in these three systems.

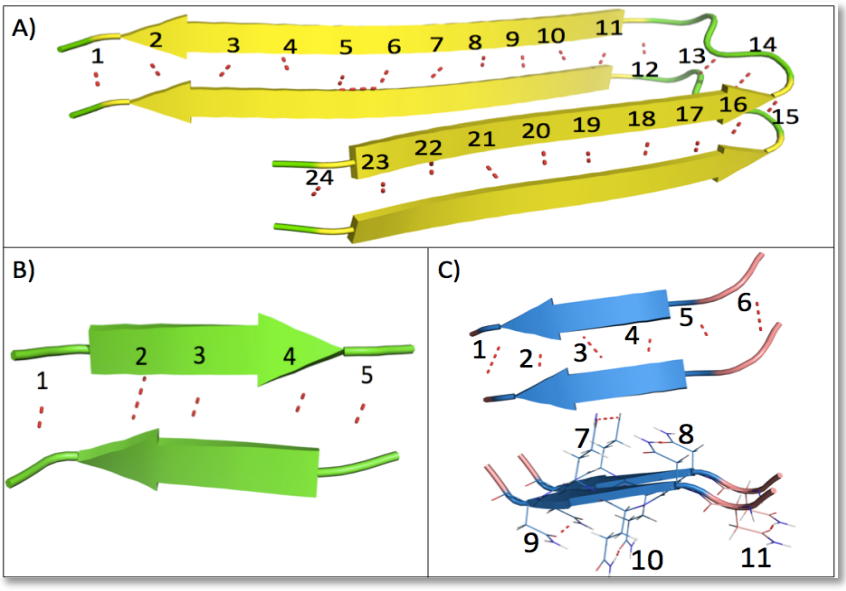


Figure 17. Effective number of hydrogen bonds. (A) 24 backbone-backbone HBs are observed in Aβ17-42**,** as it is composed of 25 residues. In the MVGGVV segment from Aβ35-40, 5 HBs of this type are present **(B).** GNNQQNY, while composed of just one more residue that MVGGVV and thus, showing 6 backbone-backbone hydrogen bonds, is able to form 5 additional sidechain – sidechain HBs through its Gln and Asn residues, which sums up a total of 11 HB interactions **(C)**. We use the term *effective HB number* to account for this situation.

Our current calculations, which are not finished yet due to the huge computational time required to achieve convergence of the results, already show that in Aβ17-42 the *re-HBC* is also present, although *hy-HBC* is not likely to show up (Fig. 18). This result is important and highlights the appropriateness of choosing this system, since the β-strands in this structure are arranged in parallel, as in the case of GNNQQNY and contrary to MVGGVV, meaning that HBC (*re-HBC*) in hydrophobic segments is also present and does not depend on the strand's configuration –at least in these two experimental structures.

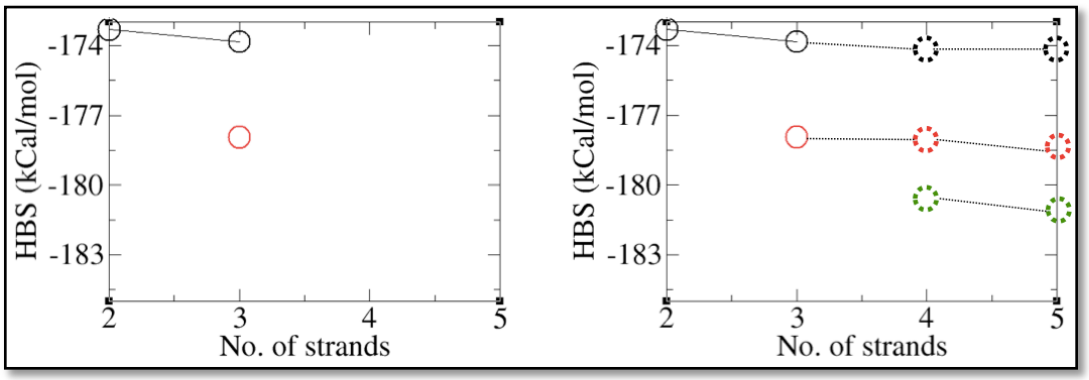


Figure 18. Cooperative effects are not restricted to polar sequences. (*Left*) Hydrogen bonding cooperativity calculations for the Aβ17-42 segment. These first results already point towards a lack of *hy-HBC* and the presence of the ubiquitous *re-HBC*. (*Right*) Extrapolation of the expected results on the basis of the three values already obtained, showing a similar trend to that seen in the left panel of **Fig. 16** for the MVGGVV segment, suggesting that *hy-HBC* is restricted to polar sequences but the residual contribution, resHBC, is always present.

Based on these results, I hypothesize that the HBC effect might arise from the sum of two contributions. The first of them corresponds to a *residual HBC (re-HBC)* due to a basal level of cooperativity present in all the amyloid-forming segments as a consequence of the highly repetitive nature of the backbone-backbone hydrogen bonds in quasi-infinite β-sheets. Considering that the torsional restrictions in extended β-strand configurations do not specifically depend on the type of residue, the amide groups in Asn & Gln side chains must be the source of a second contribution to HBC. This is proposed to be a supplementary *hyper HBC* (*hy-HBC*), which would have a negligible contribution in hydrophobic sequences. In this novel formulation, the *re-HBC* would be responsible for strengthening the HBS as the more strands dock the growing oligomer, in a sequence-independent manner. In contrast, the *hy-HBC* would provide an additional stabilizing contribution and this would account for energetic differences shown in **Fig. 16**.

Modulation of Cooperative Effects – According to our observations, fibers from GNNQQNY and Aβ17-40 in water must behave very differently with regard to their polarizable properties, given that the former is rich in Q/N residues that enable *hy-HBC*, which seems to be dominant over the length of the systems. To investigate this effect, we carried out spectroscopic analysis with broadband dielectric spectroscopy (DS). As it is shown in **Fig. 19**, fibers from GNNQQNY have a conductivity component (σ) that is about two orders of magnitude higher than those from Aβ17-40. According to our hypothesis, this difference can be attributed to the extra *hy-HBC* contribution to the overall cooperativity in GNNQQNY.

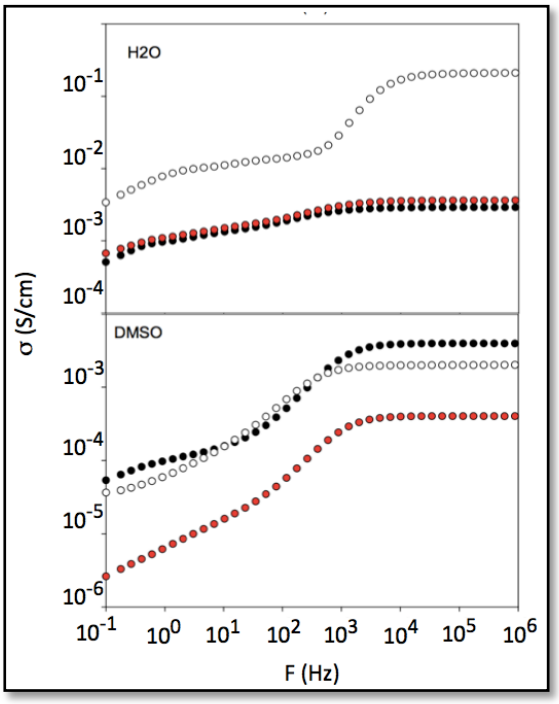


Figure 19. Dielectric Spectroscopy of fibrils from GNNQQNY and Aβ in H_2O and DMSO. (*Top*) BBDS spectra of GNNQQNY (white circles) and Aβ17-42(black ones) in H_2O. The solvent is represented as red circles. *hy-HBC* is expected to be only present in GNNQQNY and therefore, a higher hyperpolarization is expected for these fibrils. As a consequence, the conductivity component σ, obtained as the plateau of these curves, is two orders of magnitude higher in GNNQQNY than in Aβ17-42, where only *re-HBC* is present. (*Bottom*) When the DS spectra are acquired for fibrils in DMSO, the *hy-HBC* contribution in the Q/N-rich peptide vanishes and both peptides display a similar value of σ since only *re-HBC* is expected to be present, as explained in the text.

Considering the special role of the side chain amide groups of Gln and Asn, it seems plausible that choosing an appropriate molecule that can efficiently accept, but not donate H-bonds can alter the *hy-HBC* properties. DMSO is a perfect candidate in this regard, as it is a very strong HB acceptor capable of unfolding globular proteins due to its ability to form hydrogen bonds with the carbonyl groups of the peptide bonds that are stronger than the native ones [36]. We dissolved aged, preformed fibrils in this solvent and acquired their DS spectra (**Fig. 19**). Interestingly, the conductivity component in this situation is drastically reduced, consistent with a sharply decreased *hy-HBC* effect imposed by the interaction of Asn/Gln exposed NH_2 moieties with DMSO. In fact, the magnitude of s now is comparable to that of Aβ17-42, in agreement with both peptides showing only *re-HBC*.

Aged fibrils resist the denaturizing effect of 100% DMSO – To corroborate the presence of amyloid aggregates in DMSO, we carried out spectroscopic analyses with thioflavin-T (ThT) fluorescence assays on samples of GNNQQNY and Aβ17-42 which had formed amyloid fibrils in aqueous solution. These aggregates were isolated, dried and then transferred to DMSO. After one day of incubation in DMSO and centrifugation to force the fibers into the pellet and to isolate possible soluble oligomers and/or monomeric species, the ability of supernatant aliquots to enhance Thioflavin T (ThT) was measured. ThT is a fluorophore that enhances its fluorescence emission upon binding to amyloids. These assays revealed that soluble amyloid-like species from both the two different peptides were still present, which confirmed the great strength of the hydrogen bonds. This experiment was repeated after six days with the same results (Fig. 20a), which corroborates the perseverance over time of these amyloid oligomers. Therefore, whereas DMSO is an appropriate solvent to maintain Aβ monomeric [37], here we show that once the fibers are well formed and have reached the maximum cooperative effects, they are resistant to the effect of 100% DMSO.

Following these seven days of incubation in DMSO, we performed complementary solution NMR experiments, on the basis that fibers and high MW species are undetectable by this technique, whereas any monomers in solution will be readily detected. In the case of Aβ17-42, no signal was observed in the ^{1}H spectrum compared to the case of "fresh" dissolved monomer (**Fig. 20b**). This is indicative of large oligomers or suspended fibrils (as detected by ThT). In contrast, in the case of the GNNQQNY segment of Sup35, monomers are definitely present, as revealed by 1D 1H spectra (data not shown) and it was even possible to acquire a 2D ^{1}H-^{1}H TOCSY spectrum (**Fig. 20c**). In summary, although DMSO was not able to denaturalize mature Aβ17-40 fibrils, it did slowly dissociate those formed by GNNQQNY. Interestingly, when the GNNQQNY sample was transferred to D2O, its amide 1H quickly exchanged with solvent deuterons. This is consistent with a lack of peptide H-bonds which act to impede H/D exchange in protein secondary structures. (**Fig. 20c**).

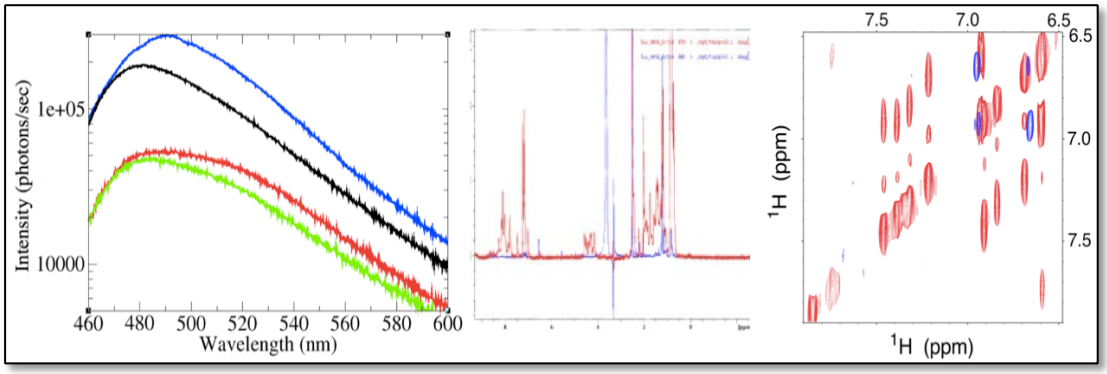


Figure 20. DMSO affects Q/N-rich fibrils but not hydrophobic ones. A. ThT assays showed that amyloid oligomers persist in the presence of 100% DMSO once the fibrils had formed (A). (**B**) 1D ^{1}H NMR spectra of Aβ freshly dissolved in DMSOd6 (red) and of Aβ fibrils incubated in DMSOd6 (blue). The lack of Aβ signals in the blue spectra means that DMSOd6 has not dissociated the fibrils. (**C**) 2D 1H TOCSY NMR spectra of GNNQQNY fibrils after incubation in DMSOd6 (red), showing the region of the spectra where side chain amide H2N and Tyr Hδ-Hε crosspeaks appear. The observation of these peaks indicates the presence of dissolved monomeric GNNQQNY. Upon addition of D2O, the signals of the side chain amide H2N moieties vanish as they quickly exchange with deuterium. The remaining signals are due to the non-exchanging Tyr Hδ-Hε.

6.2.3 Discussion and ongoing work

Our results strongly suggest that the view of HBC as arising from Q/N sidechains, and not present in hydrophobic amyloids is outdated. Here, we show that H-bonds in amyloid fibrils exhibit a general cooperative behavior, regardless of their polarity or residue composition, which we call residual hydrogen bond cooperativity. This kind of hydrogen bond cooperativity is not weakened by DMSO. Polar amyloids containing Q/N possess an additional kind of hydrogen bond cooperativity, called hyper-HBC, which can be modulated by powerful H-bonding acceptors such as DMSO. This finding was foreshadowed by a pionnering paper by Raines and Eberhardt, who suggested that HBs from Asn/Gln sidechains to mainchain may contribute more to protein stability than does a main chain – main chain one [38].

The novel theoretical description of the HBC achieved here can be reconciled on the basis of our experimental observations. When aged Ab fibers are dissolved in DMSO, oligomer dissociation is impeded, probably because DMSO is not nonpolar enough to disrupt hydrophobic interactions in Aβ17-42 amyloid. It might be interesting to test whether a more hydrophobic denaturant such as propanol, would destabilize Aβ17-42 amyloid much more strongly than the GNNQQNY fibrils. In the case of GNNQQNY, the loss of the stability from *hy-HBC* and DMSO's ability to form very strong H-bonds with HN moieties, suggests that DMSO might slowly dissociate peptides one by one from the end of the amyloid.

The proposal here that the importance of different classes of stabilizing interactions may vary between hydrophobic and polar amyloids suggests that in the future, it might be possible to develop drugs that selectively modulate the stability of one type of amyloid. This is important considering the need to inhibit pathological amyloids, like Aβ, without perturbing the activity of functional amyloids, such as the CPEB proteins implicated in memory consolidation, which have a more polar composition.

This work is being currently expanded with the inclusion of more systems in our calculations, such as the islet amyloid polypepide, also known as amylin, which is implicated in type II diabetes, and possibly using other destabilizing solvents with H-bonding properties.

6.4 References

(1) Kauzmann, W. Some factors in protein folding. *Adv. Protein Chem.* **1959**, *14*, 1-63.

(2) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133-7155.

(3) Zhang, Y. J.; Caulfield, T. Xu, Y. F.; Gendron, T. F.; Hubbard, J.; Stetler, C.; Sasaguri, H.; Whitelaw, E. C.; Cai, S.; Lee, W. C.; Petrucelli, L. The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet.* **2013**, *22*, 3112-22.

(4) Romano, V.; Quadri, Z.; Baralle, F. E.; Buratti, E. Th estructural integrity of TDP-43 N-terminus is required for efficient agrégate entrapment and consequent loss of protein function. *Prion* **2015**, *9*, 1-9.

(5) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+, a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **2009,** *44*, 213-233.

(6) Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **2004**, *278*, 353-378.

(7) Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008**, *37*, 375-397.

(8) Zhao, Z. C.; Chook, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12017-12021.

(9) Niu, C.; Zhang, J.; Gao, F.; Yang, L.; Jia, M.; Zhu, H.; Gong, W. FUS-NLS/Transportin 1 complex structure provides insights into the nuclear targeting mechanism of FUS and the implications in ALS. *PLoS One* **2012**, *7*, e7056.

(10) Qin, H.; Lim, L. Z.; Wei, Y.; Song, J. TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18619-18624.

(11) Chang, C. K.; Wu, T. H.; Wu, C. Y.; Chiang, M. H.; Toh, E. K.; Hsu, Y. C.; Lin, K. F.; Liao, Y. H.; Huang, T. H.; Huang, J. J. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem Biophys. Res. Commun.* **2012**, *425*, 219-224.

(12) Güttler, T.; Madl, T.; Neumann, P.; Deichsel, D.; Corsini, L.; Monecke, T.; Ficner, R.; Sattler, M.; Görlich, D. NES consensos redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1367-1377.

(13) Zhang, C.; Ma, J. Folding helical proteins in explicit solvent using dihedral biased tempering. *Proc. Natl. Acad. Sci.* **2012**, *109*, 8139-8144.

(14) Wang, Y. T.; Kuo, P. H.; Chiang, C. H.; Liang, J. R.; Chen, Y. R.; Wang, S.; Shen, J. C.; Yuan, H. S. The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* **2013**, *288*, 9049-9057.

(15) Romano, G.; Klima, R.; Buratti, E.; Verstreken, P.; Baralle, F. E.; Feiguin, F. Chronological requirements of TDP-43 function in synaptic organization and locomotive control. *Neurobiology of Disease* **2014**, *71*, 95-109.

(16) Fersht, A. R.; Shi, J.-P.; Jack, K.-J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Wayer, M. M. Y.; Winter, G. Hydrogen bonding and

biological specificity analyzed by proteína engineering. *Nature* **1985**, *314*, 235-238.
(17) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; Sevcik, J.; Urbanikova, L.; Myers, J. K.; Kazufumi, T.; Hebert, E. J.; Shirley, B. A.; Grimsley, G. R. Contribution of hydrogen bonds to protein stability. *Prot. Science* **2014,** *23*, 652–661.
(18) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. Evidence for resonance-assisted hydrogen bonding from cristal-structure correlations on the enol form of the .beta.-diketone fragment. *J. Am. Chem. Soc.* **1989**, *111*, 1023-1028.
(19) Bertolasi, V.; Gilli, P.; Ferretti, V.; Gilli, G. Evidence for resonance-assisted hydrogen bonding. 2. Intercorrelation between crystal structure and spectroscopic parameters in eight intramolecularly hydrogen bonded 1,3-diaryl-1,3- propanedione enols**.** *J. Am. Chem. Soc.* **1991**, *113*, 4917-4925.
(20) Cleland, W. W.; Kreevoy, M. M. Low-barrier hydrogen bonds and enzymic catalysis. *Science* **1994**, *264*, 1887-1890.
(21) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* **1994**, *264*, 1927-1930.
(22) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. Cooperativity in amide hydrogen bonding chains: implications for protein-folding models. *J. Am. Chem. Soc.* **2001**, *123*, 4348-4349.
(23) Wieczorek, R.; Dannenberg, J. J. Hydrogen-bond cooperativity, vibrational coupling, and dependence of helix stability on changes in amino acid sequence in small 3 10-helical peptides. A density functional theory study. *J. Am. Chem. Soc.* **2003**, *125*, 14065-14071.
(24) Wieczorek, R.; Dannenberg, J. J. H-bonding cooperativity and energetics of alpha-helix formation of five 17-amino acid peptides. *J. Am. Chem. Soc.* **2003**, *125*, 8124-8129.
(25) Plumley, J. A.; Dannenberg, J. J. The importance of hydrogen bonding between the glutamine side chains to the formation of amyloid VQIVYK parallel beta-sheets: an ONIOM DFT/AM1 study. *J. Am. Chem. Soc.* **2010**, *132*, 1758-1759.
(26) Plumley, J. A.; Dannenberg, J. J. Comparison of β-sheets of capped polyalanine with those of the tau-amyloid structures VQIVYK and VQIINK. A density functional theory study. *J. Phys. Chem. B.* **2011**, *115*, 10560-10566.
(27) Tsemekhman, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. Cooperative hydrogen bonding in amyloid formation. Protein Sci. 2007, 116, 761-764.
(28) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J.; McFarlane, H. T.; Madsen, A.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. Nature 2007, 447, 453-457.
(29) Miller, J. S.; Kennedy, R. J.; Kemp, D. S. Solubilized, spaced polyalanines: a context-free system for determining amino acid alpha-helix propensities. *J. Am. Chem. Soc.* **2002**, *124*, 945-962.
(30) Wu, Y.-D.; Zhao, Y.-L. A theoretical study on the origin of cooperativity in the formation of 3_{10}- and a-helices. *J. Am. Chem. Soc.* **2001**, *123*, 5313
(31) Rosseti, G.; Magistrato, A.; Pastore, A.; Carloni, P. Hydrogen bonding cooperativity in polyQ β-sheets from first principle calculations. *J. Chem. Theory Comput.* **2010,** *6,* 1777-1782.
(32) Perrin, C. L. Are short, low-barrier hydrogen bonds unusually strong? *Acc. Chem. Res.* **2010**, *43*, 1550-1557.

(33) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. Cooperative hydrogen-bonding in models of anti-parallel beta-sheets. *J. Phys. Chem. A.* **2004**, *108*, 9205-9212.
(34) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross-beta spine of amyloid-like fibrils. Nature 2005, 435, 773–778.
(35) Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Döbeli, H.; Schubert, D.; Riek, R..3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17342-17347.
(36) Jackson, M.; Mantsch, H. M. Beware or proteins in DMSO. *Biochim. Biophys. Acta* **1991**, *1078*, 231-235.
(37) Laurents, D. V.; Pantoja-Uceda, D.; López, L. C.; Carrodeguas, J. A.; Mompeán, M.; Jiménez, M. Á.; Sancho, J. DMSO affects Ab_{1-40}'s conformation and interactions with aggregation inhibitors as revealed by NMR. *RSC Adv.*, **2015**, *5*, 69761-69764.
(38) Eberhardt, E. S.; Rains, R. T. Amide-Amide and Amide-Water Hydrogen Bonds: Implications for Protein Folding and Stability. *J. Am. Chem. Soc.* **1994**, *116,* 2149-2150.
(39) Lukavsky, P. J.; Daujotyte, D.; Tollervey, J. R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F. E.; Damberger, F. F.; Allain, F. H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43 *Nat. Struct. Mol. Biol.* **2013**, *12,* 1443-1449.
(40) Kuo, P. H.; Doudeva, L. G.; Wang, Y. T.; Shen, C. K.; Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **2009**, *37*, 1799-1808.
(41) Kuo, P. H.; Chiang, C. H.; Wang, Y. T.; Doudeva, L. G.; Shen, C. K.; Yuan, H. S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **2014**, *42*, 4712-4722.
(42) Mompeán, M.; Romano, V.; Pantoja-Uceda, D.; Stuani, C.; Baralle, F. E.; Buratti, E.; Laurents, D. V. High resolution structural, stability and dynamics elucidation of the N-terminal domain of TDP-43 reveals a stable, distinct fold. *Submitted*.

Chapter 7

Conclusions

1. We have proposed a computational model for amyloid nucleation of the Sup35 heptapeptide, which shows that the rate-limiting step is the formation of a three stranded β-sheet.

2. The formation of this cross-β spine is not mediated by meta-stable water nanowires, as was previously proposed in the literature.

3. Although crystal structures of amyloid fibrils show well-ordered, perfectly straight hydrogen-bonding networks, twisted β-sheets are more likely to be present in absence of crystal packing forces. This twist slightly changes the hydrogen-bonding pattern, preserves the cooperative nature of these interactions, and makes them stronger with respect to the straight orientation.

4. We have shown that the Q/N-rich region 341-366 from the C-terminal region of TDP-43 initially adopts a chiefly disordered conformation in solution, and gradually converts into a β-rich conformation that enables oligomerization. We proposed a model for this segment in which residues 341-357 fold into a β-hairpin, followed by a well-defined region composed of residues 358-366.

5. The aggregates formed by the assembly of 341-357 β-hairpins possess many amyloid-like characteristics, as revealed by their recognition via conformational antibodies, and selective molecular probes, as well as their characterization via X-ray diffraction, and ssNMR and CD spectroscopies.

6. TEM images of TDP-43(341-357) fibrils reveal a topology similar to that formed by polyglutamine repeats. On the basis of the experimental data, computational methods were applied to build and verify a structural model in which TDP-43(341-357) β-hairpins assemble into a amyloid-like oligomer with a novel parallel, β-turn topology .

7. We used NMR spectroscopy to obtain the structure of the N-terminal domain of TDP-43 at high resolution on the basis of 60 torsion angle, 26 H-bond, and 1058 unambiguous NOE-derived distance restraints. We found that this domain adopts a stable, rare fold made up by six β-strands and an α-helix. This structure is similar to the DIX domain which mediates protein/protein interactions in Axin1, a protein involved in Wnt signaling.

8. Two Cys residues and two clusters of negatively charged residues are positioned on the surface of the NTD of TDP-43, where they may mediate interactions with other subunits or proteins.

Conclusiones

1. Se ha propuesto un modelo computacional para la nucleación del heptapétido amiloide de la proteína Sup35, mostrando que la etapa limitante es la formación de una lámina β de tres hebras.

2. La formación de la columna β-cruzada (cross-β spine) no requiere la mediación de nano-cables de agua meta-estables, tal y como se ha propuesto previamente en la bibliografía.

3. Aunque las estructuras cristalinas de fibras amiloides muestran una red de enlaces de hidrógeno ordenadas y perfectamente alineadas, en ausencia de fuerzas de empaquetamiento debidas a la cristalización es más probable que las láminas β estén curvadas. Esta torsión varía sutilmente el patrón de enlaces de hidrógeno manteniendo su carácter cooperativo, al mismo tiempo que estabiliza y hace más fuertes estas interacciones con respecto a la orientación alineada.

4. Se ha propuesto que el segmento 341-366 en la región rica en Q y N del dominio C-terminal de TDP-43 adopta inicialmente una conformación desestructurada en disolución, y que una transición gradual a una conformación rica en estructura β permite la oligomerización. Se ha propuesto un modelo para este segmento en el cual los residuos 341-357 se pliegan formando una horquilla beta, seguida de una región bien definida compuesta por los residuos 358-366.

5. Los agregados formados por las horquillas beta 341-357 tienen muchas características propias de amiloides, como se ha visto mediante experimentos de unión a anticuerpos y otras sondas moleculares selectivas por conformaciones de tipo amiloide, así como por difracción de rayos X y espectroscopias de RMN de estado sólido y de dicroísmo circular.

6. Las imágenes de microscopía electrónica de transmisión de fibras de TDP-43(341-357) revelan una topología similar a las de poliglutaminas. De acuerdo a los datos experimentales, se ha propuesto un modelo en el que las horquillas β de TDP-43(341-357) se ensamblan en una topología novedosa de giro-β paralela.

7. Se ha usado la espectroscopia de RMN para obtener la estructura de alta resolución del dominio N-terminal (DNT) de TDP-43 en base a restricciones de 60 ángulos de torsión, 26 enlaces de hidrógeno y 1058 distancias inequívocas derivadas de picos NOE. Este dominio adopta un plegamiento estable y poco conocido, compuesto por seis hebras beta y una hélice alfa. Esta estructura es muy similar al dominio DIX responsable de mediar interacciones proteína/proteína en Axin1, una proteína implicada en señalización Wnt.

8. La estructural del DNT de TDP-43 posee dos cisteínas y una agrupación de residuos cargados negativamente en la superficie que podrían mediar interacciones con otras subunidades u otras proteínas.

APPENDICES

Appendix A

Assignment of TDP-43 (342-366), Wild Type Sequence (5 ºC, pH 6.54)†

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	4.09 **57.09**	3.88 **63.62**			
Q343	‡	4.39 **55.6 ± 0.4**	2.10, 1.99 **29.2 ± 0.2**	2.38 **33.5**		
Q344	8.73	4.32 **55.6 ± 0.4**	2.08, 1.98 **29.2 ± 0.2**	2.38 **33.5**		
N345	8.67	4.70 **53.1 ± 0.2**	2.84, 2.80 **38.4**			
Q346	8.60	4.40 **55.7**	2.17, 1.99 **29.2 ± 0.2**	2.38 **33.5**		
S347	8.52	4.51 **58.3**	3.89 **63.6**			
G348	8.38	4.16, 4.12 **44.5**				
P349	-- --	4.47 **63.4**	2.31, 1.92 **32.0**	2.03 **27.2**	3.65 **49.7**	
S350	8.65	4.46 **58.4**	3.89 **63.2**			
G351	8.49	3.98 **45.2**				
N352	8.45	4.74 **53.0**	2.85, 2.76 **38.4**			
N353	8.60	4.68 **53.1 ± 0.3**	2.87, 2.78 **38.4**			
Q354	8.52	 **55.6 ± 0.4**	2.16, 2.10 **29.2 ± 0.2**	2.35 **33.5**		
N355	8.59	4.70 **53.1 ± 0.3**	2.88, 2.78 **38.4**			
Q356	8.50	4.31 **55.6 ± 0.4**	2.10, 2.00 **29.2 ± 0.2**	2.36 **33.5**		
G357	8.57	3.94 **45.2**				
N358	8.41	4.71 **53.0**	2.85, 2.77 **38.4**			
M359	8.45	4.46 **55.2**	2.10, 2.00 **32.2**	2.61, 2.52 **31.8**		2.09 **16.7**
Q360	8.50	4.31 **55.6 ± 0.4**	2.10, 1.99 **29.2 ± 0.2**	2.37 **33.5**		
R361	8.49	4.34 **55.5 ± 0.3**	1.84, 1.74 **30.9**	1.62 **27.0**	3.20 **43.0**	7.27
E362	8.64	4.58 **54.2**	2.09, 1.90 **29.2**	2.32 **35.6**		
P363	-- --	4.39 **63.4**	2.30, 1.92 **32.0**	2.03 **27.2**	3.82, 3.74 **50.4**	
N364	8.69	4.62 **53.2**	2.83, 2.80 **38.4**			
Q365	8.33	4.33 **55.6 ± 0.4**	2.13, 1.99 **29.2 ± 0.2**	2.38 **33.5**		
A366	8.23	4.10 **53.7**	1.34 **19.6**			

† In all the tables, values in **bold** are ^{13}C nuclei. The uncertainties are ±0.007 ppm for ^{1}H, ±0.1 ppm for ^{13}C and ^{15}N, unless otherwise indicated. ‡ Not observed due to fast HN <-> H_2O exchange.

Appendix B

Assignment of TDP-43 (342-366), Q343R (5 ºC, pH 6.38)*

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	4.15 **56.93**	3.98 **63.0**			
R343	‡	4.37 **56.3**	1.86, 1.79 **30.6**	1.65 **26.92**	3.22 **43.0**	7.29
Q344	8.72 ***121.2***	4.32 **55.6**	2.09, 1.99 **29.1**	2.38 **33.5**		
N345	8.66 ***119.3***	4.69 **53.1**	2.84, 2.79 **53.1**			
Q346	8.56 ***120.3***	4.40 **55.7**	2.17, 2.00 **29.1 ± 0.2**	2.37 **33.5 ± 0.2**		
S347	8.52 ***116.4***	4.51 **58.2**	3.88 **63.6**			
G348	8.39 ***109.8***	4.17, 4.13 **44.5**				
P349	-- --	4.47 **63.3**	2.31, 1.98 **32.1**	2.04 **27.2**	3.65 **49.6**	
S350	8.65 ***115.4***	4.47 **58.3**	3.92, 3.89 **63.5**			
G351	8.48 ***109.7***	3.98 **45.1**				
N352	8.44 ***117.7***	4.74 **52.9**	2.85, 2.76 **38.4**			
N353	8.60 ***118.1***	4.69 **53.1**	2.87, 2.78 **38.4**			
Q354	8.52 ***119.3***	4.31 **55.6 ± 0.4**	2.15, 2.10 **29.1 ± 0.2**	2.35 **33.5**		
N355	8.58 ***118.4***	4.70 **53.10**	2.88, 2.78 **38.4**			
Q356	8.50 ***119.7***	4.31 **55.8**	2.15, 2.00 **29.1 ± 0.2**	2.35 **33.5**		
G357	8.57 ***108.7***	3.94 **45.2**				
N358	8.41 ***117.7***	4.71 **53.1**	2.85, 2.77 **38.4**			
M359	8.45 ***119.8***	4.47 **55.3**	2.11, 2.02 **32.3**	2.61, 2.52 **31.8**		2.08 **16.6**
Q360	8.50 ***120.8***	4.33 **55.5**	2.16, 2.00 **29.2 ± 0.2**	2.36 **33.5**		
R361	8.48 ***122.1***	4.35 **55.9**	1.84, 1.74 **30.9**	1.62 **26.9**	3.20 **43.0**	7.27
E362	8.64 ***123.0***	4.58 **54.2**	2.10, 1.89 **29.2**	2.33 **35.7**		
P363	-- --	4.39 **63.3**	2.30, 1.92 **31.9**	2.04 **27.2**	3.82, 3.74 **50.5**	
N364	8.69 ***117.2***	4.62 **53.2**	2.83, 2.80 **38.4**			
Q365	8.33 ***120.3***	4.33 **55.6**	2.14, 1.99 **29.1 ± 0.2**	2.38 **33.5**		
A366	8.23 ***131.0***	4.10 **53.7**	1.34 **19.6**			

* In this table, the signals in ***bold italics*** are ^{15}N nuclei. ‡ Not observed due to fast HN <-> H_2O exchange.

Appendix C

Assignment of TDP-43 (342-366) N345K (5 ºC, pH 6.58)

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	3.94 **57.4**	3.89 **64.1**			
Q343	‡	4.39 **55.5 ± 0.2**	2.14, 2.00 **29.2**	2.37± 0.03 **33.6 ± 0.2**		
Q344	8.72	4.31 **56.0**	2.08, 1.98 **29.2**	2.39 **33.55 ± 0.2**		
K345	8.64	4.30 **56.3 ± 0.3**	1.83, 1.78 **32.9**	1.44 **24.6**	1.69 **28.8**	2.99 **41.6**
Q346	8.70	4.40 **55.8 ± 0.2**	2.12, 2.01 **29.2**	2.40 **33.55 ± 0.2**		
S347	8.64	4.51 **58.1**	3.88 **63.7**			
G348	8.44	4.18, 4.13 **44.4**				
P349	-- --	4.47 **63.2**	2.30, 1.98 **32.1**	2.03 **27.1**	3.65 **49.6**	
S350	8.67	4.46 **58.3**	3.91, 3.88 **63.5**			
G351	8.51	3.98 **45.1**				
N352	8.46	4.74 **52.9**	2.85, 2.76 **38.4**			
N353	8.60	4.69 **53.3**	2.88, 2.78 **38.4 ± 0.2**			
Q354	8.53	4.30 **56.0 ± 0.3**	2.15, 2.10 **28.9 ± 0.2**	2.37 **33.6**		
N355	8.59	4.70 **53.15**	2.85, 2.78 **38.4± 0.2**			
Q356	8.52	4.31 **56.0 ± 0.3**	2.15, 1.99 **28.9**	2.36 **33.6**		
G357	8.57	3.94 **45.2**				
N358	8.42	4.71 **53.1**	2.85, 2.77 **38.4 ± 0.2**			
M359	8.46	4.47 **55.3**	2.11, 2.02 **32.3**	2.61, 2.52 **31.8**		2.09 **16.6**
Q360	8.50	4.32 **55.6**	2.10, 2.01 **29.2**	2.37 **33.6 ± 0.2**		
R361	8.49	4.36 **55.6**	1.80, 1.74 **30.9**	1.63 **26.9**	3.20 **43.0**	7.27
E362	8.65	4.58 **54.2**	2.09, 1.89 **29.3**	2.33 **35.8**		
P363	-- --	4.39 **63.3**	2.30, 1.91 **32.0**	2.04 **27.2**	3.82, 3.74 **50.5**	
N364	8.69	4.62 **53.2**	2.83, 2.80 **38.4 ± 0.2**			
Q365	8.33	4.33 **55.5 ± 0.2**	2.14, 1.99 **29.2**	2.38 **33.6 ± 0.2**		
A366	8.24	4.10 **53.7**	1.34 **19.6**			

‡ Not observed due to fast HN <-> H_2O exchange.

Appendix D

Assignment of TDP-43 (342-366) G348V (5º C, pH 6.58)

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	4.16 **57.0**	3.99 **62.9**			
Q343	‡	4.39 **55.3 ± 0.3**	2.14, 2.00 **29.15± 0.3**	2.38 **33.5 ± 0.4**		
Q344	8.72	4.32 **55.7**	2.08, 1.99 **29.2 ± 0.3**	2.38 **33.5 ± 0.4**		
N345	8.66	4.69 **53.1 ± 0.3**	2.84, 2.79 **38.4± 0.2**			
Q346	8.57	4.38 **55.7 ± 0.3**	2.15, 1.98 **29.2 ± 0.3**	2.38 **33.5 ± 0.4**		
S347	8.54	4.55 **58.2**	3.85 **63.5**			
V348	8.33	4.47 **59.7**	2.10 **32.5**	0.99, 0.94 **20.8, 20.0**		
P349	-- --	4.45 **63.2**	2.33, 1.93 **32.0**	2.06, 1.99 **27.3**	3.90, 3.70 **50.9**	
S350	8.63	4.43 **58.2**	3.93, 3.88 **63.6**			
G351	8.58	3.99 **45.1**				
N352	8.48	4.74 **53.0**	2.85, 2.76 **38.4**			
N353	8.61	4.69 **53.1**	2.86 2.80 **38.4**			
Q354	8.52	4.30 **55.9 ± 0.4**	2.15, 2.00 **29.2 ± 0.3**	2.36 **33.5 ± 0.4**		
N355	8.59	4.70 **53.1± 0.3**	2.88, 2.79 **38.4**			
Q356	8.51	4.31 **55.9 ± 0.4**	2.16, 2.00 **29.2**	2.36 **33.5 ± 0.4**		
G357	8.57	3.94 **45.2**				
N358	8.41	4.71 **53.0**	2.86, 2.77 **38.4**			
M359	8.45	4.47 **55.3**	2.11, 2.02 **32.2**	2.61, 2.52 **31.8**		2.08 **16.6**
Q360	8.50	4.31 **55.9**	2.18, 2.01 **29.2± 0.3**	2.37 **33.5 ± 0.4**		
R361	8.48	4.35 **55.7± 0.4**	1.83, 1.74 **30.9**	1.63 **26.9**	3.20 **43.0**	7.27
E362	8.64	4.57 **54.2**	2.08, 1.89 **29.2**	2.33 **35.7**		
P363	-- --	4.40 **63.3**	2.30, 1.92 **32.0**	2.04 **27.3**	3.82, 3.74 **50.5**	
N364	8.69	4.62 **53.2**	2.85, 2.81 **38.4**			
Q365	8.34	4.33 **55.7 ± 0.4**	2.15, 1.99 **29.2**	2.38 **33.5 ± 0.4**		
A366	8.23	4.11 **53.7**	1.34 **19.6**			

‡ Not observed due to fast HN <-> H_2O exchange.

Appendix E

Assignment of TDP-43 (342-366) G348V (15 ºC, pH 6.58)

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	4.45	3.90			
Q343	‡					
Q344	8.644	4.32	2.08, 1.99			
N345	8.584	4.69				
Q346	8.49	4.38				
S347	8.68	4.45	3.85			
V348	8.237	4.47	2.09	0.98, 0.93		
P349	-- --	4.45	2.32, 1.95	2.06	3.89, 3.69	
S350	8.539	4.44	3.90			
G351	8.504	3.98				
N352	8.413	4.72	2.85, 2.75			
N353	8.536	4.68	2.87, 2.79			
Q354	8.453	4.31	2.15, 1.99	2.36		
N355	8.526	4.70	2.88, 2.77			
Q356	8.442	4.31	2.15, 1.99	2.36		
G357	8.502	3.94				
N358	8.356	4.71	2.86, 2.76			
M359	8.381	4.46	2.11, 2.02	2.61, 2.52		
Q360	8.428	4.32	2.00	2.36		
R361	8.399	4.35	1.83, 1.74	1.63	3.20	7.24
E362	8.568	4.578	2.09, 1.89	2.32		
P363	-- --	4.40	2.29, 1.93	2.03	3.81, 3.73	
N364	8.598	4.63	2.83, 2.79			
Q365	8.269	4.39	1.98	2.37		
A366	8.138	4.11	1.34			

‡ Not observed due to fast HN <-> H_2O exchange.

Appendix F

Assignment of TDP-43 (342-366) G348V (25 ºC, pH 6.58)

Residue	HN	alpha	beta	gamma	delta	epsilon
S342	‡	4.45	3.92			
Q343	‡					
Q344	8.558	4.32	2.09, 1.99			
N345	‡					
Q346	8.411	4.39				
S347	8.404	4.46	3.85			
V348	8.146	4.484	2.09	0.98, 0.93		
P349	-- --	4.45	2.32, 1.96	2.06	3.88, 3.70	
S350	8.440	4.447	3.87			
G351	8.432	3.98				
N352	8.353	4.70				
N353	8.474	4.67	2.86, 2.74			
Q354	8.391	4.31	2.15, 1.99	2.36		
N355	8.464	4.69	2.87, 2.77			
Q356	8.375	4.82	2.14, 1.99	2.36		
G357	8.437	3.94				
N358	8.299	*	2.85, 2.76			
M359	8.329	4.46	2.11, 2.02	2.60, 2.52		
Q360	8.365	4.33	2.17, 2.00	2.36		
R361	8.327	4.36	1.84, 1.74	1.62	3.20	7.21
E362	8.49	4.57	2.08, 1.89	2.32		
P363	-- --	4.40	2.30, 1.93	2.03	3.81, 3.73	
N364	8.511	4.64	2.84, 2.80			
Q365	8.198	4.39	2.14, 1.98	2.37		
A366	8.049	4.12	1.34			

‡ Not observed due to fast HN <-> H_2O exchange.
*missing signal in HN region and overlapped in other regions of the spectrum

Appendix G

Assignment of TDP-43 (341-357), Wild Type Sequence (5 ºC, pH 6.62)†

Residue	HN	alpha	beta	gamma	delta	epsilon
A341	‡	4.12 **51.5**	1.54 **19.2**			
S342	‡	4.45 **58.4**	3.91 **63.7**			
Q343*	‡	4.32 **55.8 ± 0.2**	2.14, 1.98 **29.3**	2.36 **33.7**		
Q344	‡	4.31 **55.9**	2.09, 1.98 **29.3**	2.36 **33.7**		
N345*	‡	4.67 **53.3 ± 0.2**	2.84, 2.76 **38.6**			
Q346	‡	4.38 **55.7**	2.17, 1.99 **29.2**	2.37 **33.7**		
S347	‡	4.49 **58.2**	3.87 **63.7**			
G348	‡	4.13 **44.5**				
P349	-- --	4.46 **63.5**	2.30, 1.97 **32.1**	2.03 **27.2**	3.65 **49.7**	
S350	‡	4.49 **58.2**	3.87 **63.7**			
G351	‡	3.96 **45.2**				
N352	‡	4.73 **53.0**	2.84, 2.76 **38.6**			
N353*	‡	4.67 **53.3 ± 0.2**	2.84, 2.76 **38.6**			
Q354*	‡	4.32 **55.8 ± 0.2**	2.14, 1.98 **29.3**	2.36 **33.7**		
N355*	‡	4.67 **53.3 ± 0.2**	2.84, 2.76 **38.6**			
Q356*	‡	4.32 **55.8 ± 0.2**	2.14, 1.98 **29.3**	2.36 **33.7**		
G357	‡	3.74 **45.9**				

† As in the other tables, values in **bold** are ^{13}C nuclei. The uncertainties are ±0.007 ppm for ^{1}H, ±0.1 ppm for ^{13}C and ^{15}N, unless otherwise indicated. ‡ Not observed due to fast HN <-> D_2O exchange. *,* Inter-residual ambiguities.

Appendix H

Secondary Structure from Analysis of Circular Dichroism Spectra of TDP-43 peptides at 25 °C, pH 6.8*

Algorithm: SELCON3, protein reference set 4.

	Q343R	N345K	G348V	WT fresh	WT aged
helix	0.08	0.07	0.12	0.12	0.07
β-sheet	0.24	0.20	0.29	0.23	0.43
turn	0.25	0.24	0.23	0.22	0.20
coil	0.44	0.43	0.36	0.43	0.30
RMSD	0.21	0.28	0.25	0.32	0.57

Algorithm: SELCON3, protein reference set 7.

	Q343R	N345K	G348V	WT fresh	WT aged
helix	0.13	0.12	0.11	0.09	0.09
β-sheet	0.33	0.40	0.19	0.15	0.47
turn	0.12	0.17	0.09	0.12	0.23
coil	0.40	0.17	0.65	0.64	0.22
RMSD	0.12	1.46	0.05	34	0.47

Algorithm: CDSSTR, protein reference set 4.

	Q343R	N345K	G348V	WT fresh	WT aged
helix	0.06	0.06	0.07	0.09	0.05
β-sheet	0.32	0.30	0.32	0.31	0.44
turn	0.25	0.26	0.25	0.25	0.21
coil	0.35	0.36	0.36	0.35	0.30
RMSD	0.04	0.03	0.03	0.02	0.01

Algorithm: CDSSTR, protein reference set 7.

	Q343R	N345K	G348V	WT fresh	WT aged
helix	0.02	0.01	0.03	0.04	0.04
β-sheet	0.28	0.32	0.18	0.15	0.44
turn	0.16	0.20	0.12	0.12	0.25
coil	0.52	0.45	0.66	0.68	0.28
RMSD	0.04	0.03	0.03	0.02	0.01

*Do to rounding off, these fractional values may not always add up to 1.00.

Appendix H (cont.)

Secondary Structure from Analysis of Circular Dichroism Spectra of TDP-43 peptides at 25 ºC, pH 6.8*

Algorithms: K2D & CONTINLL (protein reference set 7).

Algorithm: K2D	WT freshly dissolved	WT aged	**Algorithm: CONTINLL (ref set 7)**	WT fresh	WT aged
helix	0.08	0.03	helix	0.07	0.05
β-sheet	0.43	0.50	β-sheet	0.13	0.38
This method does not consider turns.			turn	0.09	0.27
coil	0.49	0.47	coil	0.70	0.31
RMSD	0.097	>0.22	RMSD	0.05	0.20

Mean for all Algorithms ± σ	WT freshly dissolved	WT aged
helix	0.08 ± 0.03	0.06 ± 0.02
β-sheet	0.23 ± 0.12	0.44 ± 0.04
‡turn	0.16 ± 0.07	0.23 ± 0.03
‡coil	0.56 ± 0.16	0.28.± 0.04

*Do to rounding off, these fractional values may not always add up to 1.00.

‡The values from the K2D method, which lumps turn and coil together, were not included in the calculation of the mean turn and coil values.

Appendix I

NMR Spectral Parameters.

A. 13C, 15N NTD sample in 15% D_2O, 85% H_2O pH 3.8. Spectra for backbone & sidechain assignments.

Spectra #/Pulse Sequence	Ns	Matrix	Spectral Width (ppm)	Comments
801,802,830,832,840,843,848, 849, 853 1D ^{1}H	8-64	32k	13	Redox state of TCEP, Detect -OH
803, 2D ^{1}H-^{15}N HSQC	4	2k x 128	12.0 x 40	
804, 807, 810, 817, 831, 839, 842,850/ 2D ^{1}H-^{15}N HSQC	4	2k x 128	12.0 x 15.8	Folded to increase resolution; control spectra to Monitor State of Sample,
805 ^{1}H-^{13}C HSQC	8	1k x 256	14 x 80	Centered to reveal aliphatic 13C
806 3D hnco	8	2k x 48 x 64	12 x 15.8 x 16	Sequential
808 3D hncai	8	2kx 48 x 64	12 x 16 x 30	assignment
809 3D hncoca	8	2k x 48 x 64	12 x 15.8 x 30	
811 cbcacohn	8	2k x 48 x 80	12 x 15.8 x 75	
812 hncacb	16	2k x 48 x 80	12 x 15.8 x 75	
814 BECE 2D ^{1}H-^{15}N HSQC	16	2k x 480	12 x 15.8	^{13}C-filtered based on side chain
815 GACE 2D ^{1}H-^{15}N HSQC	64	2k x 60	12 x 15.8	^{13}C-filtered based on side chain
816 BECEi2 2D ^{1}H-^{15}N HSQC	8	2k x 480	12 x 15.8	^{13}C-filtered based on side chain
817 proplus 2D ^{1}H-^{15}N HSQC	64	2k x 80	12 x 15.8	Reveals residues following Pro
818 prominus 2D ^{1}H-^{15}N HSQC	64	2k x 80	12 x 15.8	Reveals residues preceeding Pro
830,832 1D 1H				
834 3D hcccohhgpwg3d3	8	2k x 48 x 80	14 x 15.8 x 75	Side chain assignment
835 3D hcccohhgpwg2d3	8	2k x 48 x 80	14 x 15.8 x 75	Side chain assignment
836 2D ^{1}H-^{13}C HSQC	8	1k x 256	14 x 80	Centered to reveal aromatic ^{13}C
837 3D hcchdigp3d	8	2k x 48 x 90	14 x 70 x 14	TOCSY/1H side chains
838 3D hcchdigp3d2	8	2k x 48 x 90	14 x 70 x 70	TOCSY/13C side chains
840 3D NOESY- 13C-1H HSQC	16	2k x 64 x 90	14 x 70 x 14	Aliphatic groups
841 3D NOESY- 13C-1H HSQC	16	2k x 64 x 90	14 x 40 x 14	Aromatic groups
851 3D hcccohn3d2	8	2k x 4 x 410	12 x 15.8 x 6	Side chain assigning
852 3D hbhacohn3d	8	2k x 4 x 410	12 x 15.8 x 6	HA, HB assigning
854 (4,2)D hnnh3d1	8	2k x 4 x 310	12 x 32 x 4.85	HN-HNi assigning
855 (4,2)D hnnh3d2	8	2k x 4 x 204	12 x 32 x 4.85	HN-HNi assigning
858 1lechot2	64		13	Echo sequence for estimation of the correlation time
859 1lechot2	64		13	

Appendix I (cont.)

NMR Spectral Parameters

B. 13C, 15N NTD sample in 15% D_2O, 85% H_2O pH 3.8. Spectra to measure temperature coefficients.

Spectra #/Pulse Sequence	Ns	Matrix	Spectral Width (ppm)	Comments
819 1D, 820 2D 15N HSQC	4	2k x 128	12.0 x 40	25ºC
821 1D, 822 2D 15N HSQC	4	2k x 128	12.0 x 40	20ºC
824 1D, 825 2D 15N HSQC	4	2k x 128	12.0 x 40	15ºC
826 2D 13C HSQC	4	2k x 128	12.0 x 40	15ºC
827 1D, 828 2D 15N HSQC	4	2k x 128	12.0 x 40	30ºC
829 1D, 839 2D 15N HSQC	4	2k x 128	12.0 x 40	34ºC
831 1D, 832 2D 15N HSQC	4	2k x 128	12.0 x 40	39ºC
833 1D, 834 2D 15N HSQC	4	2k x 128	12.0 x 40	5ºC
835 1D, 836 2D 15N HSQC	4	2k x 128	12.0 x 40	10ºC

Appendix I (cont.)

NMR Spectral Parameters

C. Unlabeled NTD sample in 100% D_2O pH 3.88.

Spectra #/Pulse Sequence	Ns	Matrix	Spectral Width (ppm)	Comments
870 1D 1H	64	32k	13	
872 1D 1H	64	32k	13	002 h 10 m H/D Ex
873 1lechot2	64	32k	13	Echo sequence for estimation of the correlation time
874 1lechot2	64	32k	13	
875/2D NOESY	80	2k x 512	12 x 12	Mixing time = 80 ms
876 1D 1H	64			016 h 25 m H/D Ex
877 2D 1H-13C HSQC	80	1k x 128	12 x 80	Natural abund. 13C
878 1D 1H	64	2k x 512	12 x 12	026 h 25 m H/D Ex
908 1D 1H	64	2k x 512	12 x 12	048 h 35 m H/D Ex
909 1D 1H	64	2k x 512	12 x 12	120 h 50 m H/D Ex
910 1D 1H	64	2k x 512	12 x 12	168 h 7 m H/D Ex
911 1D 1H	64	2k x 512	12 x 12	168 h 12 m H/D Ex
912 1D 1H	128	2k x 512	12 x 12	288 h 25 m H/D Ex
914 1D 1H	128	2k x 512	12 x 12	360 h 0 m H/D Ex
915 1D 1H	128	2k x 512	12 x 12	360 h 5 m H/D Ex

Appendix I (cont.)

NMR Spectral Parameters

D. 13C, 15N NTD sample in 15% D_2O, 85% H_2O pH 3.8. Spectra for elucidation of backbone dynamics.

Spectra #/Pulse Sequence	Ns	Matrix	Spectral Width (ppm)	Comments
8831, 8832, 8833, 8834, 8845, 8836, 8837, 8838, 8839, 8840 2D $^{1}H-^{15}N$ HSQC [hsqct1etf3gpsi3d]	4	2k x 128	12.0 x 15.8	Folded to increase resolution; Experiments for determination of T1.
885, 886, 887, 888, 889, 890, 891, 892, 893, 894 2D $^{1}H-^{15}N$ HSQC [hsqct2etf3gpsi]	4	2k x 128	12.0 x 15.8	Folded to increase resolution; Experiments for determination of T2.
896, 897, 898, 899, 900, 901, 902, 903, 904, 905 2D $^{1}H-^{15}N$ HSQC [hsqct2etf3gpsi]	4	2k x 128	12.0 x 15.8	Folded to increase resolution; Experiments for determination of T1Rho.